DEVELOPMENT OF ENFORCED STACKED INTERCALATORS THAT TARGET TRINUCLEOTIDE REPEAT MISMATCHES IN DNA

BY

JULIO FRANCISCO SERRANO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Steven C. Zimmerman, Chair
Professor Raven Huang
Professor Douglas A. Mitchell
Professor Wilfred A. van der Donk

ABSTRACT

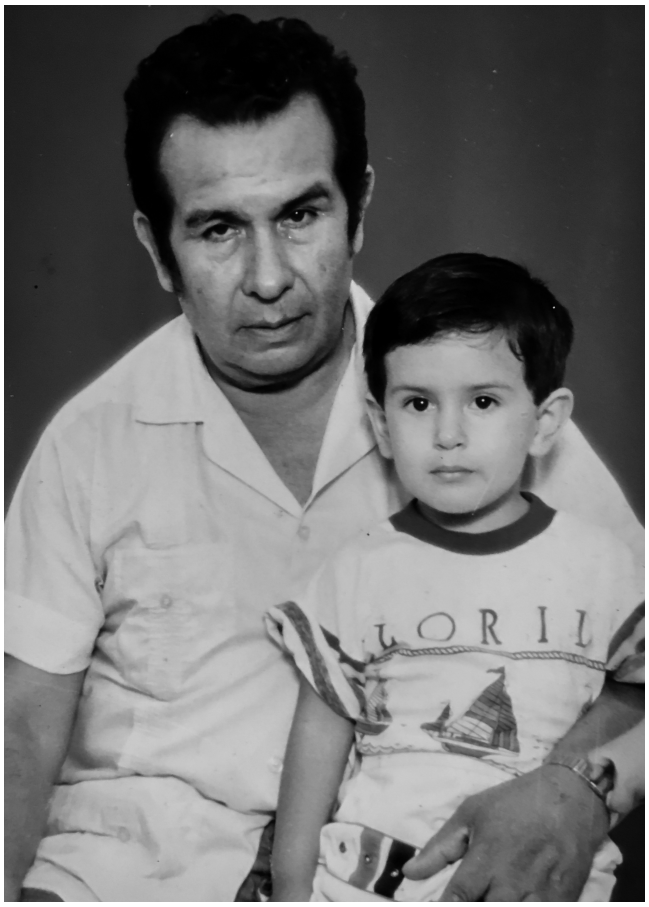
Mismatched base pairs are ubiquitous in more than 30 hereditary disorders whose origin can be traced to unstable repeating sequences in genomic DNA. For instance, non-Watson–Crick T–T mismatch base pairs flanked by G–C base pairs appear in myotonic dystrophy type 1 (DM1), which is caused by the expansion of CTG trinucleotide repeats (TNR) in the 3'-untranslated region of the dystrophia myotonica protein kinase (DMPK) gene. Extensive efforts have made it possible to elucidate its pathogenesis and mechanism of action. In this regard, numerous compounds have been developed that target the toxic repeating CUG RNA transcript ($r(CUG)^{exp}$) and inhibit its sequestration of key pre-mRNA splicing proteins, such as MBNL1. Indeed, since we reported our first ligand **JFA** in 2009, our group has been at the forefront in discovering ligands that specifically bind U–U mismatches and can potentially be used as therapeutics. More recently, we have started to focus our efforts to discover molecules that target the parent CTG expanded repeats ($d(CTG)^{exp}$) in hopes to develop a more promising therapeutic. A brief overview of targeting DNA, DNA mismatches, DM1 and our ligands is given in Chapter 1.

A previous anisotropy screen in our lab of ligands that inhibit the MBNL1–$r(CUG)^{exp}$ complex identified a novel lead pyrroloquinazoline compound. However, as we will see in Chapter 2, further investigation revealed that it was selective for T–T mismatches. Interestingly, we noted that mode of binding toward CTG mismatched sites appeared to be cooperative. Due to its limited water solubility, we attempted to develop further derivatives. We took into consideration these observations to develop and improve our acridine-base ligands in the last chapter.

To further improve the affinity and selectivity of our acridine-based ligands, we proceeded to develop a series of enforced stacked intercalators. Our first approach involved utilizing a larger mismatch-recognition diaminopurine unit, as the described in the first part of Chapter 3. However, although selective for T–T and C–C mismatches, its nonspecific binding was not improved. Finally, we employed a macrocyclic design, and developed a small library of macrocycles. We showed that these ligands selectively bind to CTG trinucleotide repeats in DNA with negligible nonspecific binding to duplex DNA. In addition, they were about twice as effective and selective in inhibiting transcription than the control. Lastly, we discovered that the macrocyclic structure design of our ligands did not necessarily correlate with a reduction

in their cytotoxicity in HeLa cells, as we had previously hypothesized. The synthesis, biophysical studies, and *in vitro* activity of these macrocyclic ligands are described in the second part of Chapter 3.

Sin ninguna duda,
Esto es para ti, Betty,
Para ti, Tío Will,
Y para ti, Mamá Marina.



(circa 1993)

En memoria de mi Papá Pancho

ACKNOWLEDGEMENTS

My work presented here and, most importantly, my development as a scientist, would never have been possible without my friends, family, and mentors. My time in graduate school has taught me beyond anything I could have ever hoped — scientifically, physically, and psychologically. I have questioned myself countless of times, but I truly believe it was the people who were there and whom I have met that made it possible to finish this journey.

I must first express my deepest gratitude to my advisor, Professor Steven Zimmerman. I am forever indebted for your professional and generous guidance, endless understanding of my mistakes and faults, and continuous encouragement on my struggles from the very first day. I always admired your work ethic and eye for details, exceptional ability to guide through problems by asking questions that lead to an answer, and your constant optimism and focus on the overall purpose: "What is the big picture?" — something that I strive continuing to develop every day. I know it was not always easy to be my advisor, but I know that I would not be half the scientist I am now if you were not there to push me.

No doubt that the lab and 348 RAL became my second home, and my lab mates became my extended family. I am especially thankful for past and present DM subgroup members including Brenda Andrade, JuYeon Lee, Lien Nguyen, Long Luu, Lauren Hagler, Ke Li, and Iti Kapoor. I am also very thankful for other Z-Group members, including Yugang Bai, Junfeng Chen, Dawn Ernenwein, Edzna García Ramírez, Ephraim Morado, Shampa Samanta, Gary Suseno, Gretchen Vincil, Brittany Walker, Jonathan Wang, Justin Wright, Thao Xiong, and Henry Xu. The group has changed a lot in the past 5 years, and no doubt that the atmosphere is a lot more vibrant. Hoping for many more international potlucks in the future! Thanks also to my undergraduate students Jane Phang and Dan Curet for their valuable contribution in my projects and for making me laugh constantly. I am very lucky to have had the privilege of being your mentor.

Finally, I would like to thank the core people who were my rocks and key support during these past five years. First, I would like to thank my uncle Will for being a father figure and for always pushing me to be charismatic, optimistic, to have a laugh no matter the situation, and never forget my roots. I will always remember the talks you would give me when I was a kid, and for supporting me and being very open with anything about which I would talk with

you. Whenever I am back home with you, I overflow with joy and laughter, and that is the best therapy. It has been a long way since I first worked with you on the weekends and summers, from the age of 12 until the weekend before I moved to Illinois.

Second, I thank my wife Kali for her unconditional love and support, for always attempting to show me when enough is enough, and for being patient with me whenever I get hyper for no reason. I do not exaggerate when I say I would not have finished many projects or deadlines if you had not been there to help me think clearer and push me till the end. I truly feel like the luckiest man in the world, and meeting you (in the Z-Group!) was the best thing that happened to me in graduate school (expect maybe for that one acridine compound…). I cannot wait to grow old with you and spend the rest of our lives together.

Lastly and absolutely most importantly, I thank my mom, Betty. You once told me you learned to be a mother with me and you aren't perfect, but you have no idea how perfect of a mom I think you are to me. I thank you for sacrificing more than words can describe. You are the strongest woman I know; for you there is nothing I wouldn't do. Always accepting my mistakes with a smile on your face and instilling in me to never be satisfied with good enough. You always believed in me, and your selflessness is unmatched. Every time that I talk to you, I learn something new. I hope to one day be a fraction of what you are.

RECONOCIMIENTOS

Mi trabajo presentado aquí y, lo más importante, mi desarrollo como científico, nunca hubiera sido posible sin mis amigos, mi familia, y mis mentores. Mi tiempo en la escuela de posgrado me ha enseñado más de lo podría haberme imaginado — científicamente, físicamente y psicológicamente. Me he cuestionado a mí mismo más de lo que jamás hubiera imaginado, pero realmente creo que fueron las personas que estuvieron allí y a quienes conocí que hicieron posible terminar este camino.

Antes que nada, debo expresar mi más profunda gratitud a mi mentor, Profesor Steven Zimmerman. Estaré siempre endeudado por su guía profesional y generosa, comprensión sin fin de mis errores y faltas, y ánimo continuo en mis problemas desde mi primer día. Siempre he admirado su ética de trabajo y su ojo por los detalles, su optimismo constante y su enfoque en el objetivo principal: "¿Cuál es la idea general?" — algo que continúo desarrollando todos los días. Yo sé que no fue fácil ser mi mentor, pero sé que yo no sería la mitad del científico que soy hoy si no fuese por usted.

No hay duda de que el laboratorio y 348 RAL se convirtieron en mi segundo hogar, y mis compañeros se convirtieron en mi familia extendida. Estoy especialmente agradecido por los miembros y ex miembros del sub-grupo DM, incluyendo Brenda Andrade, JuYeon Lee, Lien Nguyen, Long Luu, Lauren Hagler, Ke Li, and Iti Kapoor. También estoy agradecido por los otros miembros del Grupo Z, incluyendo Yugang Bai, Junfeng Chen, Dawn Ernenwein, Edzna García Ramírez, Ephraim Morado, Shampa Samanta, Gary Suseno, Gretchen Vincil, Brittany Walker, Jonathan Wang, Justin Wright, Thao Xiong, and Henry Xu. El grupo ha cambiado bastante en los últimos 5 años, y no hay duda de que el ambiente es más vibrante. ¡Ojalá sigan las fiestas de comida internacionales en el futuro! También agradezco a mis aprendices de pregrado Jane Phang y Dan Curet por sus contribuciones valorables en mis proyectos y haciéndome reír constantemente. Soy muy afortunado de haber tenido el privilegio de ser su mentor.

Finalmente, me gustaría gradecer a la gente central quienes fueron mis rocas y apoyo clave en estos últimos cinco años. Primero, agradezco a mi Tío Will por ser como mi papá y siempre empujándome a ser carismático, optimista, a reírme sin importar la situación, y nunca olvidar mis raíces. Siempre recordaré tus charlas cuando era un chico, y tu apoyo y franqueza con

cualquiera de mis preguntas. Cada vez que regreso a casa, yo reboso con alegría y risa, y esa es la mejor terapia. Ha sido un largo camino desde que comencé a trabajar contigo en los fines de semana y veranos, desde los 12 años hasta el fin de semana antes de mudarme a Illinois.

Segundo, agradezco a mi esposa Kali por su amor y apoyo incondicional, por siempre tratando de mostrarme cuando ya es suficiente, y por ser tan paciente conmigo. Yo no exagero cuando digo que no hubiera terminado varios proyectos si no hubieses estado allí para ayudarme a pensar más claramente y empujarme hasta el final. De verdad que me siento como el hombre más suertudo del mundo, y haberte conocido (en el Grupo Z!) fue lo mejor que me sucedió en Illinois. No puedo esperar a envejecer contigo y pasar el resto de nuestras vidas juntos.

Por último y absolutamente lo más importante, agradezco a mi mamá, Betty. Una vez me dijiste que tú aprendiste a ser mamá conmigo y que no eras perfecta, pero no tienes idea de lo perfecta de una madre que yo pienso que eres. Te agradezco por sacrificar más de lo que las palabras pueden describir. Tú eres la mujer más fuerte que conozco; por ti no hay nada que yo no haría. Siempre aceptando mis errores con una sonrisa e inculcando en mí en nunca estar satisfecho con lo suficientemente bueno. Siempre creíste en mí, hasta en mis peores momentos, y tu generosidad es inigualable. Espero algún día ser una fracción de la persona que tú eres.

"En el camino encontraras piedras, curvas, pendientes, subidas y bajadas, pero al final veraz el sol en el horizonte. Tú siempre brillarás con luz propia."

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Targeting Nucleic Acids with Small Molecules Mismatches

DNA and RNA targeting with small molecules represents a well-established area of research and is of great academic and pharmaceutical interest. One of the major turning points in the area of DNA binding with small molecules was when Lerman hypothesized the concept of the intercalation of acridine derivatives (*e.g.*, proflavine) in DNA in 1961,[1] for which seminal crystallographic studies were reported in the late 1970s that showed DNA could elongate to allow space to "sandwich" an acridine molecule between two base pairs.[2,3] The concept of binding DNA as a therapeutic strategy was evident from the onset, inspiring research on natural product intercalators such as the powerful anticancer drug actinomycin D.[4]

Although there are many other natural products that bind DNA (*e.g.*, echinomycin[5], distamycin A[6]), synthesizing and developing small molecules in the laboratory that recognize specific nucleotide sequences has remained a challenge.[7] Indeed, the majority of small molecules bind non- or poorly specifically and through two broad major and well-characterized binding modes: intercalation or groove binding.[8,9] Intercalators typically contain an aromatic core that facilitates binding by insertion between base pairs, and more complex intercalators (Figure 1) contain further functional groups that facilitate groove binding. Notably, intercalators that are highly selective towards complex DNA structures such as triplexes[10] and quadruplexes[11] have been studied in recent years. In addition, intercalation reduces the natural DNA helical twist (*e.g.*, ethidium untwists DNA by 26º).[12] This distortion prevents intercalators from binding in back-to-back base pairs, known as the nearest-neighbor principle.[13]

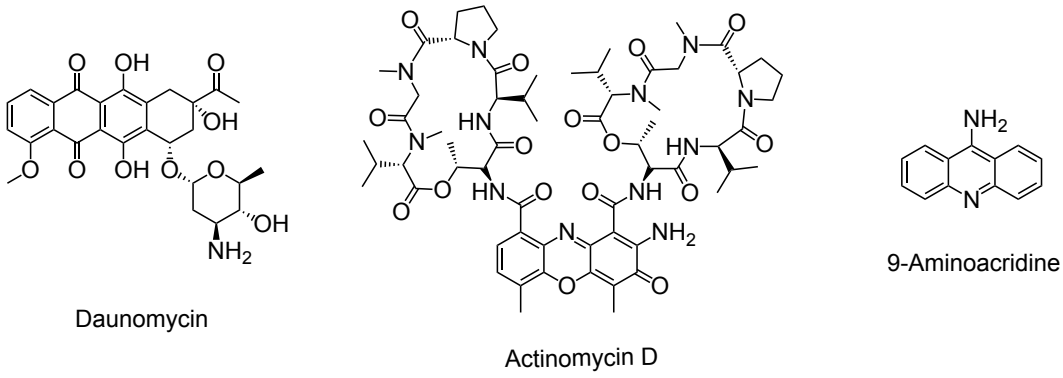


Figure 1. Examples of intercalators.

On the other hand, groove binders are typically positively-charged Instead of the fused aromatic ring architecture seen in intercalators, groove binders are typically positively charged, crescent-shaped molecules constituted of linked aromatic rings (Figure 2a).[14] Some well-known groove binders include Hoechst 33258, netropsin and berenil.[12] Perhaps the most well-known synthetic sequence-specific binders are imidazole-pyrrole ligands developed by Dervan, with affinities in the nano- to picomolar range (Figure 2b).[15-17] These groove binders

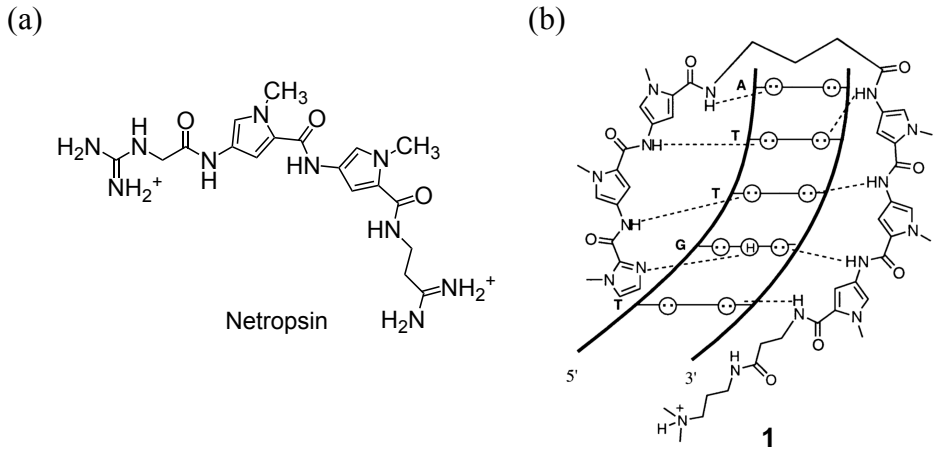


Figure 2. (a) Chemical structure of groove binder netropsin. (b) Model for the complex formed between hexapeptide Dervan-type groove binder **1** and a DNA duplex with a 5'-TGTTA-3' site. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines. Circles containing an H represent the N(2) hydrogen of guanine. Reproduced with permission from ref. 17. Copyright © 1994, American Chemical Society. All rights reserved.

were able to specifically discriminate A–T and T–A base pairs by utilizing pyrrole–hydroxypyrrole and hydroxypyrrole-pyrrole pairings, respectively.

Several small molecules that specifically target nucleic acids have been studied and shown to interfere with DNA processes such as replication, transcription, translation, and also impact cell proliferation.[9] Likewise, molecules that bind RNA exhibit similar effects such as interfering with translation by binding mRNA directly.[18] Accordingly, the design and discovery of small molecules that bind DNA and RNA defects, such as mismatches and abasic sites, in a selective and specific fashion is key to the development of DNA-targeting drugs (and the side-effects of existing drugs), but still remains an important challenge.[19] (For an excellent book on this topic, see *Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes*.[7])

1.2. Binding and Recognition of DNA Mismatches

DNA mismatches are defects that occur when two non-complementary bases are aligned in the same base-pair step of a duplex. Mismatches can appear during replication, formation of heteroduplexes during recombination, and even spontaneous deamination of nucleobases.[18] Damage to DNA bases can also be caused by physical factors (*e.g.*, UV radiation) and environmental damaging chemicals (*e.g.*, alkylating agents), and can lead to the formation of mismatched base pairs followed by mutagenesis, dysfunctions and diseases.[20] If uncorrected,

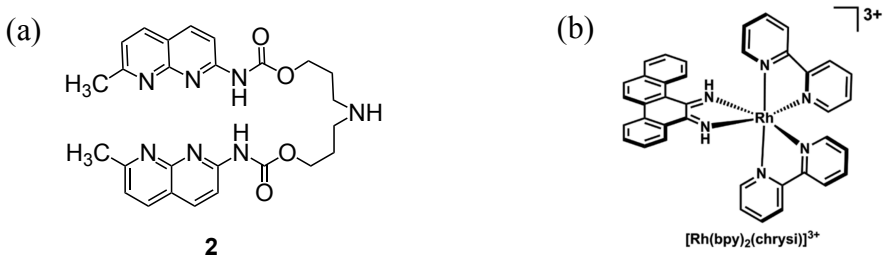


Figure 3. Chemical structure of (a) naphthyridine dimer **2** and (b) $[Rh(bpy)_2(chrysi)]^{3+}$ complex designed for recognition of the G–G mismatch.

DNA mismatches can ultimately give rise to genetic mutations and disease states.[21] Consequently, the MutS protein of the mismatch repair (MMR) system, perhaps the best-characterized DNA repair system,[22] recognizes and corrects damaged sites with excellent efficiency and specificity (up to 1000-fold higher affinity compared to normal duplex).[23]

In efforts to gain insight into natural nucleotide mismatch recognition systems, the last few decades have seen an enormous spike in the development of small molecules that site-specifically recognize mismatches, as well as the characterization of mismatches. Indeed, the structures of most mismatched base pairs have been reported (X-ray crystallography or NMR studies in solution), with interesting features such as wobble pairs, protonated bases, unusual base tautomers, and others.[24] Pioneering efforts to develop mismatch-selective ligands include the work by Barton, who developed the intercalating rhodium(III) complex[25] $[Rh(bpy)_2(chrysi)]^{3+}$ and a series of derivatives with excellent mismatch selectivity;[26,27] and also the work by Nakatani, who has developed a series of dimeric naphthyridine intercalators capable of hydrogen-bonding with mismatched nucleobases specifically in a triplet pattern due to their unique mode of binding (Figure 3).[28] In addition, as will be discussed in Chapter 3, cyclobisintercalators have also been studied to selectively bind DNA mismatches, an area pioneered by Lehn and Teulade-Fichou.[29-31]

Besides their high potential as therapeutic candidates or diagnostic tools in fields such as molecular genetics, mismatch-selective small molecules also offer potential applications to manipulate and analyze DNA *in vitro*. In essence, to achieve efficient function, these small molecules must be able to target specific mismatched sites in an "ocean" of normal Watson-Crick base pairs. Undoubtedly, Nature is once again way ahead of us.

1.3. Trinucleotide Repeat Diseases

Mismatched base pairs are ubiquitous in more than 30 hereditary disorders whose origin can be traced to unstable repeating nucleotide sequences in genomic DNA. Although the majority of expanded repeats consist of trinucleotide sequences in the form $(CXG)^{exp}$ (X = any nucleotide), there are a some with tetra-, penta-, and even dodecameric repeats.[32] DNA trinucleotide repeats are actually common within the human genome; however, it is the expansion of these repeats that alters gene expression and human cells, which leads to genetic disorders. Some of the more well-known trinucleotide repeat (TNR) expansion diseases include Fragile-X Syndrome (CGG), Huntington's disease (CAG), Friedreich's ataxia (GAA) and myotonic dystrophy (CTG).[33-35] Each disease has a unique base repeat sequence, either in a coding or noncoding region of DNA, and their location in the genome. In contrast to static mutations, which are maintained in somatic tissues and stably passed down from generation to generation, mismatch repeat mutations continue to expand in length within tissues and generations.[36] The length of the repeat tract is known to correlate with disease severity and genetic anticipation (*i.e.*, age of onset). So far, no cures or approved treatments are available for these diseases, only palliative therapeutics to treat their symptoms.

1.4. Myotonic Dystrophy

Discovered in 1992, myotonic dystrophy type 1 (DM1) is one of the two major types of myotonic dystrophy, which is the most common form of adult muscular dystrophy, and affects about 1 in 8,000 people worldwide.[37-39] As with the other TNR diseases, there are currently no cures or treatments for this disease; only medications to alleviate symptoms. Patients suffer from progressive muscle weakness, sustained muscle contractions, cardiac defects, and other neuromuscular problems.[37]

1.4.1. Disease Mechanism

DM1 is caused by the unstable expansion of CTG trinucleotide repeats (d(CTG)^{exp}) (50 to >2,000 repeats) in the 3'-untranslated region of the dystrophia myotonica protein kinase (*DMPK*) gene in chromosome *19q13* (Figure 4).[34,36,38] In fact, CTG•CAG trinucleotide repeat expansions at distinct chromosomal loci is the mutation present in several other neurological diseases such as Huntington's disease (HD), and many forms of spinocerebellar ataxia (SCA).[40,41] Like other TNR expansion diseases, repeat length correlates directly with DM1 disease severity and inversely with the age of onset – a phenomenon termed anticipation.[38]

The expansion of trinucleotide repeats is a complicated process believed to occur during the transient exposure of single DNA strands, such as during replication, recombination, transcription, and by multiple repair mechanisms.[33,34,42] Most models agree that the repetitive nature of dsDNA promote the formation of "slipped-stranded" DNA conformation through denaturation and renaturation.[35,43] In the case of DM1, it is thought that the transcription of CTG•CAG repeats, by transiently exposing single DNA strands, allows long CAG and CTG

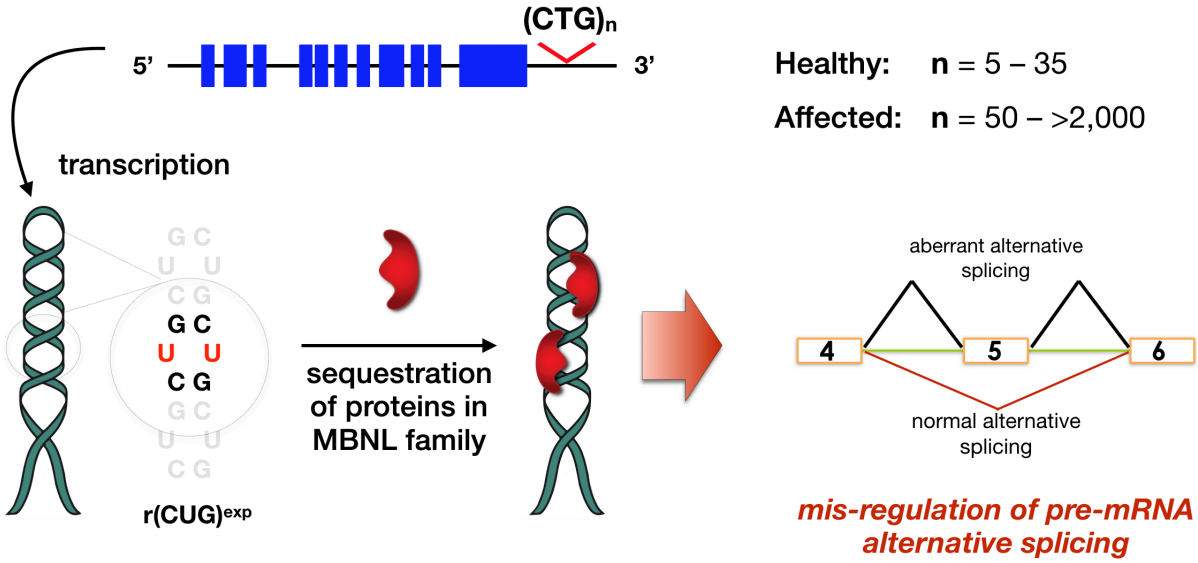


Figure 4. Schematic overview of the molecular basis for DM1.

repeat tracts to form abnormal secondary structures, such as d(CAG)^{exp} slipped-strand DNA, and d(CTG)^{exp} hairpins structures in which non-Watson–Crick T–T mismatch base pairs flanked by G–C base pairs.[42,44] The formation of these stable unusual secondary structures kinetically "trap" the expanded repeating unit, which engages in several DNA repair processes and stall various DNA polymerases.[42,44] This results in the misalignment of repetitive DNA strands, causing expansions or contractions of the repeats.[45]

Studies support a gain-of-function mechanism of the expanded RNA transcript. Upon transcription of the d(CTG)^{exp} strands, the repeating r(CUG)^{exp} folds into an extended hairpin that contains multiple copies of 5'-CUG/3'-GUC motifs, which bind muscleblind-like 1 protein (MBNL1) and form ribonuclear foci (Figure 4).[46] The r(CUG)^{exp} can also cause aberrant activation of protein kinase C (PKC), which leads to hyperphosphorylation, stabilization, and up-regulation of the CUG-binding protein 1 (CUGBP1).[47] These two proteins are coregulators of various alternative splicing transitions, and their disruption has been correlated with dysregulation of pre-mRNA splicing, some of which are known to directly result in disease symptoms such as insulin insensitivity (IR splicing), myotonia (CLCN1 splicing), and cardiac defects (cTNT splicing).[48] Likewise, it was shown that r(CUG)^{exp} disrupted the translation of the MEF2 protein, which affects multiple levels of mRNA and miRNA in heart tissues.[49] In addition, other r(CUG)^{exp}-induced toxic pathways have recently been suggested to be involved in the DM1 pathogenesis. Notably, the bidirectional transcription of the CTG•CAG repeats produce r(CAG)^{exp} and r(CUG)^{exp} transcripts,[50] which further undergo repeat-associated non-ATG (RAN) translation and make multiple toxic homopeptides.[51] Further, it was shown that r(CUG)^{exp} disrupted the translation of the MEF2 protein, which affects various levels of mRNA and microRNA in heart tissues.

1.4.2. Therapeutic Strategies for DM1

One therapeutic approach for DM1 is to disrupt the harmful interactions between MBNL1 and r(CUG)^{exp} by using small molecules. This was shown in a groundbreaking study by Swanson in which the overexpression of MBNL1 in a DM1 mouse model resulted in the reduction of myotonia and restored the normal adult-splicing patterns for 4 pre-mRNAs.[52] Consequently, Miller reported three lead compounds from a dynamic combinatorial library (DCL) screening for synthetic RNA-binding ligands that inhibit MBNL1–r(CUG)^{exp} interaction.[53] Additionally, Berglund discovered that pentamidine disrupted the MBNL1–r(CUG)^{exp} complex and reversed the missplicing of several pre-mRNAs in HeLa cells.[54]

Alternatively, another potential strategy is to target the CTG•CAG trinucleotide repeat expansions. Although a drug that targets the toxic r(CUG)^{exp} would alleviate the disease symptoms, there are some drawbacks to this approach. Importantly, 1) the r(CUG)^{exp} transcripts would be continuously produced as the parent DNA is still present, 2) the patient's expanded CTG repeats would continue to expand with time and could potentially result in needing larger quantities of therapeutics, and 3) continuation of disease progression continues further transmission between generations with a reduced age of onset and a significantly more severe form of the disease.[32,35,36]

Although multiple processes cause the expansion of TNR and their detailed mechanisms are not yet known,[36,55-57] one of the replication models for repeat instability suggests that formation of a stable secondary structure by a repetitive run in the lagging-strand template stalls lagging-strand synthesis and that repeat contractions occur if a DNA polymerase involved in repair of the gap skips the structured portion of the lagging-strand template.[34,36,43] Thus, developing a small molecule therapeutic that selectively binds hairpin d(CTG)^{exp} or

CTG•CAG triplet repeats in DNA and can inhibit its transcription or induce its contraction could potentially be a powerful strategy to cure the disease, although it would be an especially challenging goal because multiple processes cause the expansion and their detailed mechanisms are not yet known.[36,55-57] As of now, there are only a handful of reports on small molecules that have induced contractions and decreased the rate of expansion of the CTG•CAG triplet repeat,[58-60] and two reports that utilize CRISPR-Cas9 technology to contract CTG•CAG triplet repeats.[57,61]

In line with this strategy, our group and others have developed small molecules that bind r(CUG)^{exp} and inhibit the sequestration of MBNL1 (see Section 1.5 for an overview of ligands from our lab).[52-54,62-72]

1.5. Development of Small Molecules in the Zimmerman Group

In 2009, we reported a rationally-designed acridine-based intercalator (*i.e.*, ligand **JFA**) with a triaminotriazine ring as a "Janus-wedge" recognition unit, resulting in very low micromolar affinity to CTG and CUG mismatched sites in DNA and RNA, respectively.[62] This triaminotriazine unit is believed to interact with the mismatched thymine or uracil units through its two hydrogen bonding faces and allowing complementarity with the mismatches (Figure 5). Indeed, subsequent biophysical studies and molecular dynamics simulations suggested a major groove triplet binding between the mismatching nucleotides and the triaminotriazine

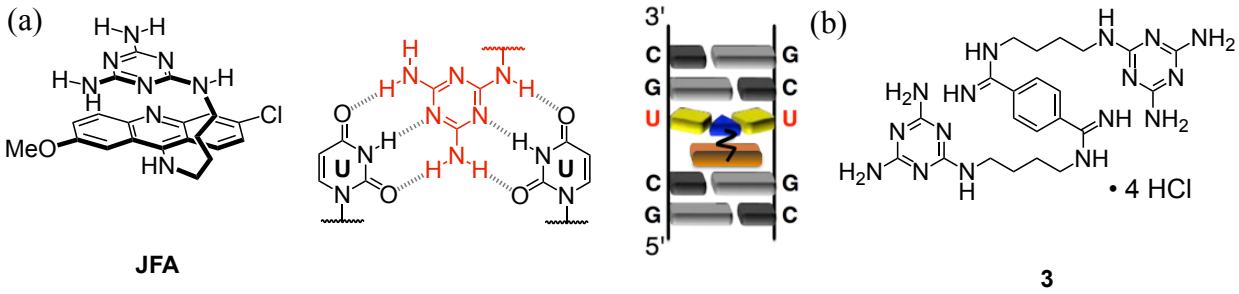


Figure 5. (a) Chemical structure of ligand **JFA** and its proposed Janus-wedge-type binding mode with mismatched base pairs (black) through hydrogen-bonding of the triaminotriazine unit (red). (b) Chemical structure of **3** with a bisamidinium groove binding motif and two triaminotriazine units.

ring, while the acridine ring ultimately intercalates between the mismatch and the neighbor Watson-Crick base pair.[73] Further derivatization[64] and development of dimeric analogs[65] was done to increase the ligand's water solubility and inhibition potency.

Due to the general cytotoxicity of intercalators and limited aqueous solubility, a second type of inhibitor was developed (ligand **3**) with improved solubility and almost no cytotoxicity.[63] This ligand contains a bisamidinium groove-binding motif, which was previously used by Butcher in 2011 for a ligand that binds to the stem-loop region of the HIV-1 frameshift site (FS) RNA, an A-form helix similar to r(CUG)^{exp}.[74] Ligand **3** also contains two triaminotriazine units to increase the affinity and selectivity by the simultaneous recognition of two CUG mismatch sites. Indeed, foci reduction was observed with as little as 5 μM of **3** in d(CTG)$_{960}$-transfected HeLa cells. Ligand **3** was also shown to correct the misregulated splicing of IR transcript. Derivatives were developed that could intervene in multiple DM1 disease pathways,[75] and dimerized through azide-alkyne "click" chemistry to produce a powerful inhibitor of the MBNL1–r(CUG)^{exp} complex.[76] In addition to finding new cores or scaffolds for the development of new novel types of inhibitors, as with our pyrroloquinazoline ligand in Chapter 2, we are also interested in to further enhancing and exploring new designs of our current lead compounds in our lab, as we will explain in Chapter 3.

CHAPTER 2

DISCOVERY OF A NEW DNA-BINDING LIGAND

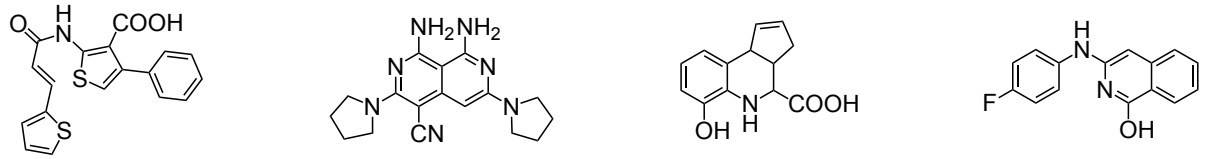
2.1. Introduction

2.1.1. High Throughput Screening to Identify Lead Compounds for DM1

One of the most widespread therapeutic approaches for DM1 has been to disrupt the harmful interactions between MBNL1 and r(CUG)^{exp} by using small molecules. Disney and coworkers reported that small molecules that bind to MBNL1 induced splicing defects and DM1 disease phenotypes.[48] Indeed, a major goal of our lab is to discover and develop low-molecular weight therapeutic agents that target the toxic r(CUG)^{exp} in DM1. Traditionally, we had utilized rational design to develop our lead compounds. We developed our first lead compound (**JFA**)[62] based on the reported crystal structure of r(CUG)$_6$ by Berglund in 2005.[77] Likewise, we developed our second-generation bisamidinium groove-binding ligand based on the reported NMR solution structure of an RNA frameshift site stimulator and its binding in the stem-loop region of the HIV-1 frameshift site (FS) RNA.[74] Although rational design provides the freedom to custom-design small molecules, it is often a slow method to survey a large chemical space and discover new potential lead compounds.

A high-throughput screen (HTS) of a library of small molecules is a powerful approach that would expedite the discovery and identification of possible lead compounds. In 2012, Austin and coworkers reported two assays that could be used orthogonally in a HTS fashion for the detection of ligands that inhibit the formation of the toxic MBNL1–r(CUG)^{exp} complex.[78] The assay utilizes biotinylated r(CUG)$_{12}$ and MBNL1-His$_6$ along with fluorescently labelled streptavidin and a fluorescently labelled anti-His$_6$ antibody. When the MBNL1–r(CUG)^{exp} complex forms, the two fluorophores form a fluorescence resonance

energy transfer (FRET) pair and time-resolved (TR)-FRET can be measured. A small molecule inhibitor of the complex would result in no TR-FRET signal. In 2013, Disney reported four lead compounds that were identified with this assay (Figure 6) and noted that **5** induced splicing defects of the IR and cTNT mini-genes towards a DM1-like phenotype, providing evidence that supports the strategy of targeting the RNA instead of the protein to inhibit the MBNL1–r(CUG)^{exp} complex.[48]



4 **5** **6** **7**

Figure 6. Small molecules that inhibit the formation of the MBNL1–r(CUG)$_{12}$ complex as identified by the Disney group through a screen of the NIH's MLPCN library.

2.1.2. Lead Compound Discovery Through High Throughput Fluorescence Anisotropy Screening

Previously in our lab, small molecule inhibitors were screened in efforts to discover and identify new lead compounds for DM1 therapeutics. Yen-Jun Ho tested the binding of several molecules to the MBNL1–r(CUG)^{exp} complex using a HTS fluorescence anisotropy assay.[79] This technique was reported by Shapiro and coworkers in 2006 to discover small molecule inhibitors of the vigilin-vitellogenin mRNA complex.[80] First described by Perrin in 1926, fluorescence polarization is based on the observation that when plane-polarized light excites stationary fluorescent molecules, these will emit light back in a fixed plane. However, fluorophores usually undergo rotational diffusion more rapidly than the time required for light emission. This results in largely randomized emission planes, effectively depolarizing most of

the emitted light. In contrast, when the fluorophore is bound to a larger molecule (*e.g.*, a protein), the larger effective molecular volume causes the rotation to be slower. Light emission is more likely to be in the same plane as that during excitation and remain highly polarized. Thus, fluorescence anisotropy measures the loss of polarization of the light emitted from the fluorophore, which is affected by the rate of tumbling of the bound complex.

The bound state should be much larger than the free state to obtain the best signal difference. Yen-Jun Ho used a r(CUG)$_6$ with TAMRA attached to the 5'-end as the RNA construct, and GST-MBNL1N (56 kDa) as the protein.[79] From over 140,000 compounds stored at the National Cancer Institute, the 1,597-compound Diversity Set III was chosen for the initial assay development based on pharmacologically desirable features and representation of novel pharmacophores, among others. The fluorescence anisotropy screen resulted in the identification of 14 hits, and following subsequent confirmatory EMSA assays for the displacement of radio-labeled r(CUG)$_{12}$ and activity in the presence of competitor tRNA narrowed this to 4 lead compounds (Figure 7).[79] Of these four compounds, ligand **11** showed the most promise and was investigated further due to its simpler synthetic scheme.

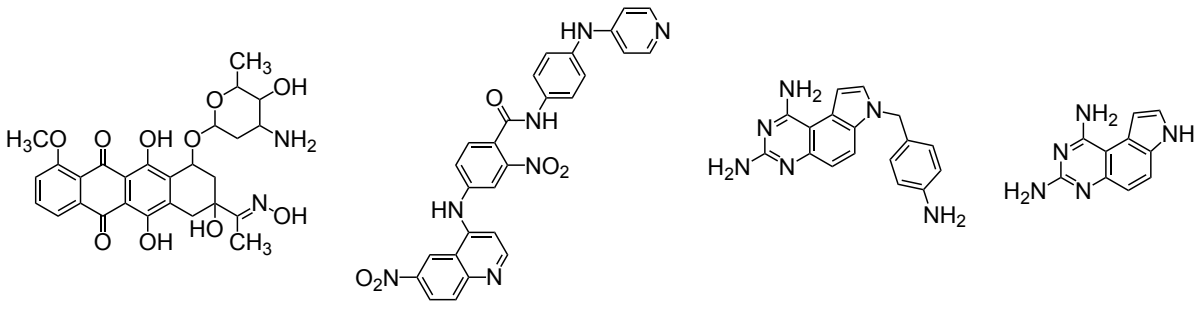


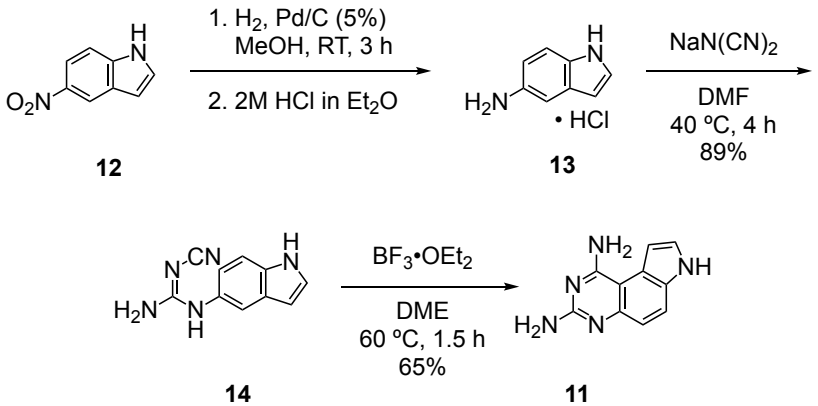
Figure 7. Lead compounds identified through a fluorescence anisotropy screen by Yen-Jun Ho.

2.1.3. Ligand 11 Inhibits the MBNL1–CUG Complex

One of the hits identified as a MBNL1–r(CUG)^{exp} complex inhibitor was the pyrroloquinazoline ligand **11** (Figure 2.2), which, along with its derivatives, had been previously described as an inhibitor of dihydrofolate reductase in 1996 by Stables[81] and assessed for antimalarial activities in 2005 by Lin.[82] When studied with TAMRA-r(CUG)$_6$ and GST-MBNL1N, ligand **11** exhibited an IC_{50} of 61 ± 15 μM with a calculated K_i of 15 ± 3 μM,[79] which were similar to other small molecule inhibitors (*e.g.*, pentamidine[54], IC_{50} = 58 ± 5 μM). These results provided preliminary promise that ligand **11** could be used as a lead compound to further derivatize and optimize as a r(CUG)^{exp} binder.

2.2. Discovery that Ligand 11 Binds d(CTG)$_2$

As mentioned in Chapter 1, developing a small molecule therapeutic that targets the parent d(CTG)^{exp} and could inhibit its transcription or lead to its contraction is another potential strategy to cure the disease, although fully detailed mechanisms are not yet known.[36,55-57] This is potentially a step towards curing the disease rather than disrupting protein-RNA interactions. Thus, our group has recently decided to screen the binding capabilities of ligand **11** against other DNA mismatches to further assess its selectivity.



Scheme 1. Synthesis of ligand **11**.

The pyrroloquinazoline ligand was synthesized from commercially available 5-aminoindole. Condensation with dicyanamide to give the cyanoguanidine derivative **14** occurred in excellent yield. Finally, ring closure to form the diaminopyrimidine ring was carried out with boron trifluoride etherate (Scheme 1). It is important to point out the structural and hydrogen-bonding similarities between ligand **11** and the triaminotriazine recognition unit used in ligand **JFA**.

To test the binding properties and selectivity of ligand **11**, isothermal titration calorimetry (ITC) experiments were performed with 10-base oligonucleotides: r(CUG)$_2$, d(CTG)$_2$, d(CAG)$_2$, and dsDNA (Table 1). ITC allows $\Delta G°$, $\Delta H°$, and $T\Delta S°$ for a ligand-receptor interaction to be determined by measuring stepwise changes in the enthalpy of interaction during the course of a titration experiment.[83] To our surprise, ligand **11** was found to bind to d(CTG)$_2$ with micromolar affinity (K_D ~ 6 μM) but *did not* exhibit measurable binding to r(CUG)$_2$ (Figure 8). Although this raised the question of how ligand **11** was able to inhibit the MBNL1–r(CUG)^{exp} complex, we decided to further investigate its DNA-binding properties

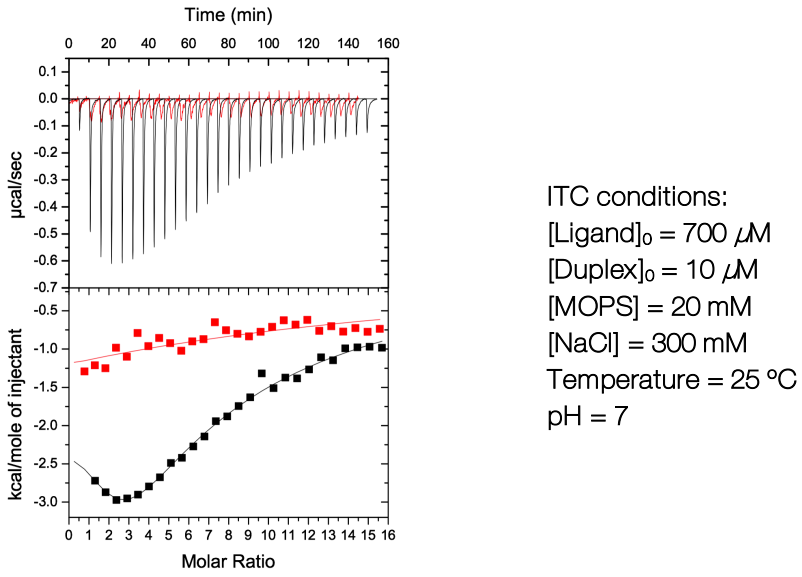


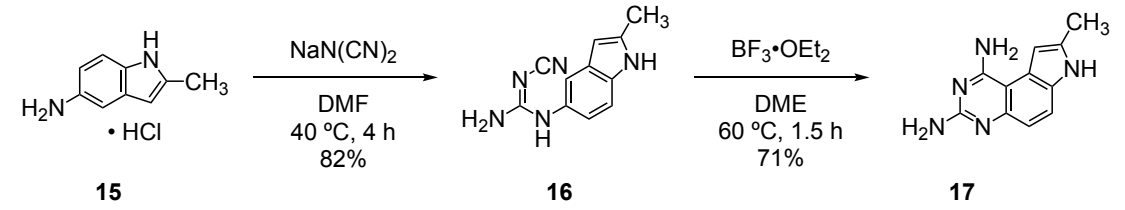
Figure 8. ITC profiles of ligand **11** with (a) d(CTG)$_2$ (black) and r(CUG)$_2$ (red).

and assess its potential as a possible lead CTG•CAG-binding compound (specifically, the d(CAG)^{exp} slipped-strands or d(CTG)^{exp} hairpins structures, and their transcription inhibition or length contraction).

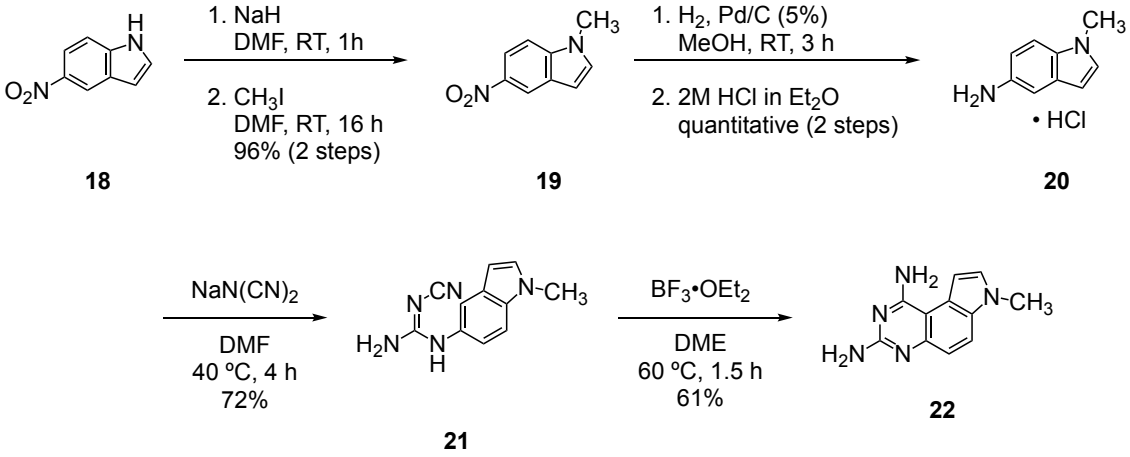
2.3. Investigation of Mode of Binding of Ligand 11 with d(CTG)$_2$

Against the other DNA sequences, ligand **11** showed weak and nonspecific binding. These preliminary results were encouraging and suggested that ligand **11** is selective of the d(CTG)$_2$ duplex. To further investigate the mode of binding and whether ligand **11** could be derivatized without affecting the binding toward d(CTG)$_2$, methylated derivatives **17** and **22** were prepared *via* a similar synthetic route as ligand **11** (Scheme 2). Their binding affinities were tested using ITC with both d(CTG)$_2$ and r(CUG)$_2$ to establish a structure-activity relationship (SAR).

(a)



(b)



Scheme 2. Synthetic schemes of methylated derivates **17** (a) and **22** (b).

Ligands **17** and **22** showed similar binding affinity profile toward $d(CTG)_2$ as ligand **11**, suggesting that the side from C(1) to N(4) might be involved in the binding, whereas the pyrrole nitrogen N(7) is not (Figure 12). Thus, ligand **17** was synthesized from commercially available 5-amino-2-methylindole (Scheme 2a), and the synthesis of ligand **22** began with N-methylation of 5-nitroindole, followed by hydrogenation with Pd/C to obtain the aminoindole, and the last two steps were the same as for ligand **11** (Scheme 2b).

The ITC experiments showed that a large quantity of ligand needed to be titrated into the instrument cell to display saturation (*i.e.*, reduction in heat released with each injection). Upon the annealing of the two ssDNAs to form a thermodynamically stable duplex structure containing one T–T mismatch, one would expect a titration point after 1 molar equivalent of ligand is titrated. Surprisingly, the titration curves display a small, but noticeable, "tear-shaped" drop as the heat released slightly increases from the first to fourth injection, followed by less heat released in subsequent injections. These characteristics of the titration isotherm appeared to show a cooperative binding mode.[84-86]

It is possible to obtain sharper transitions in ITC isotherms by increasing the concentration of macromolecule/host (*e.g.*, protein, DNA, RNA, etc.).[85] Changing the concentration alters

A	B	C	D
GC	GC	GC	GC
AT	UT	AT	AT
CG	CG	CG	CG
CG	CG	CG	CG
GC	**GC**	**GC**	**GC**
CG	**U•U**	**T•T**	**A•A**
CG	**CG**	**CG**	**CG**
GC	GC	GC	GC
CG	CG	CG	CG
GC	GC	GC	GC
5′	5′	5′	5′

Table 1. List of oligonucleotide duplexes used in biophysical studies.

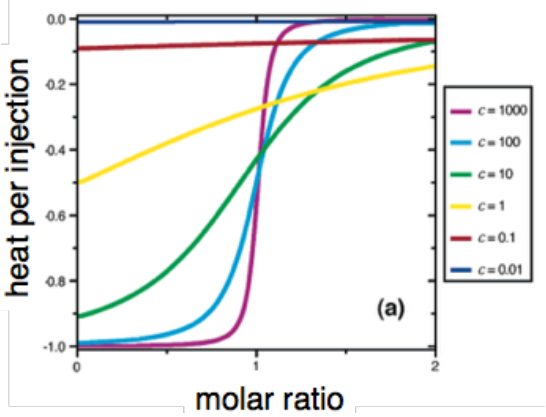


Figure 9. Plot of the effects on the binding isotherm of different ***c*** values.

the critical parameter c, known as the Wiseman constant, which determines the shape of the binding isotherm. The value $c = K_b * n * [DNA]_{tot}$ where K_b is the equilibrium binding constant and $[DNA]_{tot}$ is the total concentration of binding sites in the cell. ITC experiments are designed to have a c value of between 1 and 100 to obtain a sigmoidal isotherm allowing accurate calculation of K, n, and ΔH values. When c values are >50, the binding isotherms produce relatively sharp transitions, whereas experiments with low c values produce much shallower transitions and extend to large $[L]_{tot}$:$[DNA]_{tot}$ (Figure 9).[87] Therefore, by varying the DNA concentration, a family of related isotherms is obtained. Figure 8 shows a sharper isotherm curve as the DNA concentration is increased from 10 μM to 60 μM, each fit with a sequential binding of two sites model (provided the lowest χ^2 value). This model assumes the first ligand to bind to the macromolecule at the first site, and the nth ligand to bind goes to the nth site. The best fit will generate K, ΔS, and ΔH for each site. This model predicts that the first binding event occurs with a K_D = 6.4 ± 3 μM, and a second weaker binding event with a K_D = 350 ± 100 μM. ITC experiments with higher DNA concentrations (*i.e.*, 30 and 60 μM) would provide better fitting and allow us to obtain fuller saturation in heat of titration, but this would require higher amounts of ligand to show saturation. Since the ligand is only soluble in DMSO, this means more DMSO will be used which could inhibit the secondary structure of DNA and

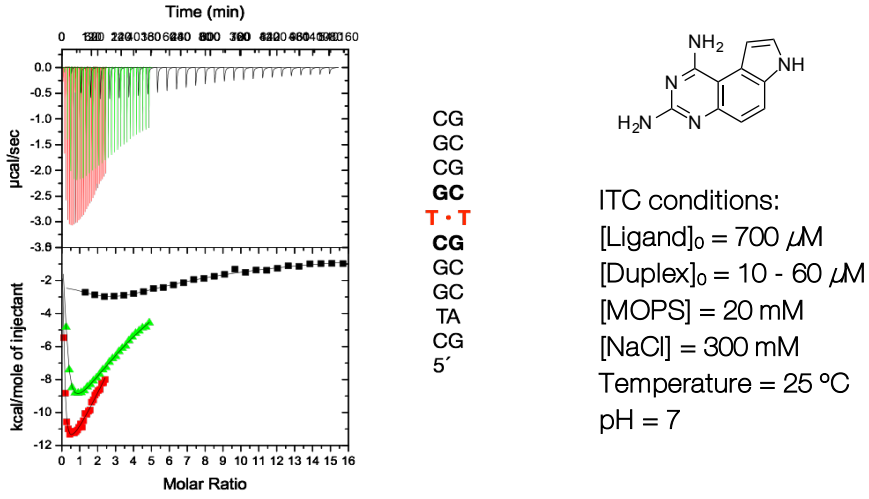


Figure 10. ITC profiles of ligand **11** with DNA concentrations of 10 μM (black), 30 μM (green), and 60 μM (red) on the same scale.

produce excess background heat of dilution throughout the titration experiment. Thus, a more water-soluble derivative is needed to saturate the higher DNA concentrations and avoid possible issues from DMSO.

In addition to ITC experiments with ligand **11**, analyses using the method of continuous variation (*i.e.*, Job plot) and thermal melting curves and were performed to further explore the binding mode of ligand **11** with d(CTG)$_2$. In the method of continuous variation, solutions of guest and host (*e.g.*, ligand **11** and d(CTG)$_2$, respectively) are combined so that the mole ratio of guest:host varies over a wide range while the sum of their concentrations (*i.e.*, total number of moles) remains constant. Thus, the maximum amount of d(CTG)$_2$:**11** will form when the d(CTG)$_2$:**11** ratio is the stoichiometric ratio. Ligand **11** and its complex with d(CTG)$_2$ show a λ_{max} at 319 nm and it is also fluorescent. This allowed us to monitor the formation and quantity of the d(CTG)$_2$:**11** complex *via* fluorescence spectroscopy (λ_{ex} = 319 nm; λ_{em} = 340–600 nm, $\lambda_{em_{max}}$ = 454 nm). The Job analysis was performed at a total concentration of 60 μM and indicated a stoichiometry of 1:1 (ligand to DNA) (Figure 9). When the concentration was

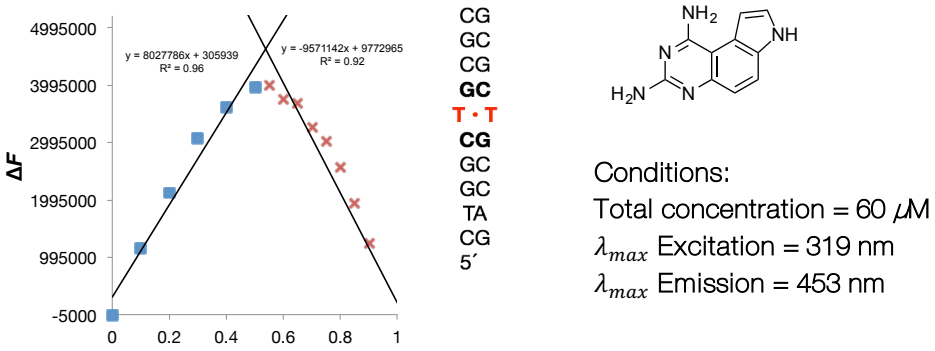


Figure 11. Job analysis of ligand **11** binding to d(CTG)$_2$ at a total concentration of 60 μM.

increased to 90 μM, a 1:1 ratio was similarly obtained. This study did not agree with the ITC experiments because the ITC binding isotherm seemed to show a higher stoichiometry of binding. ITC with higher DNA concentration provide a sharper binding isotherm, but, as of this writing, we have only been able to use up to 5 and 2.5 molar equivalents of ligand without exceeding 7.5% DMSO (Figure 8a). These results suggest the possibility of a single relatively high-affinity binding site, followed by a lower-affinity binding site not observed under the concentrations used for the Job analysis. Again, we would need a more water-soluble derivative of ligand **11** to increase its concentration.

Mismatch stabilization by ligand **11** was studied through thermal denaturation of the resulting complex. Ligand **11** was studied with d(CTG)$_2$ and dsDNA. The results show a melting temperature increase of 2.3 ºC when one equivalent of ligand **11** was used with

	T_m (ºC)	ΔT_m (ºC)		T_m (ºC)	ΔT_m (ºC)
d(CTG)$_2$	50.7		DNA duplex	69.84	
+ 1 equiv. ligand **11**	52.9	2.3	+ 1 equiv. ligand **11**	68.52	-1.32
+ 2 equiv. ligand **11**	48.8	-1.9	+ 2 equiv. ligand **11**	66.1	-3.74
+ 3 equiv. ligand **11**	48.4	-2.3	--	--	--

Table 2. Melting temperatures of d(CTG)$_2$ and fully complimentary DNA duplex with different equivalents of ligand **11**.

d(CTG)$_2$, and decreases in melting temperature when 2 and 3 equivalents were used (Figure 10). In contrast, only decreases in melting temperature were observed in thermal denaturation experiments of dsDNA and ligand **11** (Table 2).

2.4. Proposed Water-soluble Ligand 11 Derivatives

As noted earlier, methylation at the N(7) and C(8) position did not affect the ITC binding isotherm of ligand **11** significantly (Figure 12). Thus, derivatives **28** and **36** were proposed to increase the water solubility (Schemes **3** and **4**, respectively). The proposed ligand **28** incorporates a dimethylamino group at the C(8) position to increase the water solubility; this group has previously been used in our laboratory to increase the water solubility of some ligands. Further derivatization at the N(7) position has been reported to reduce solubility even in DMSO.[88] Other amines can be incorporated to the acyl chloride intermediate early in the

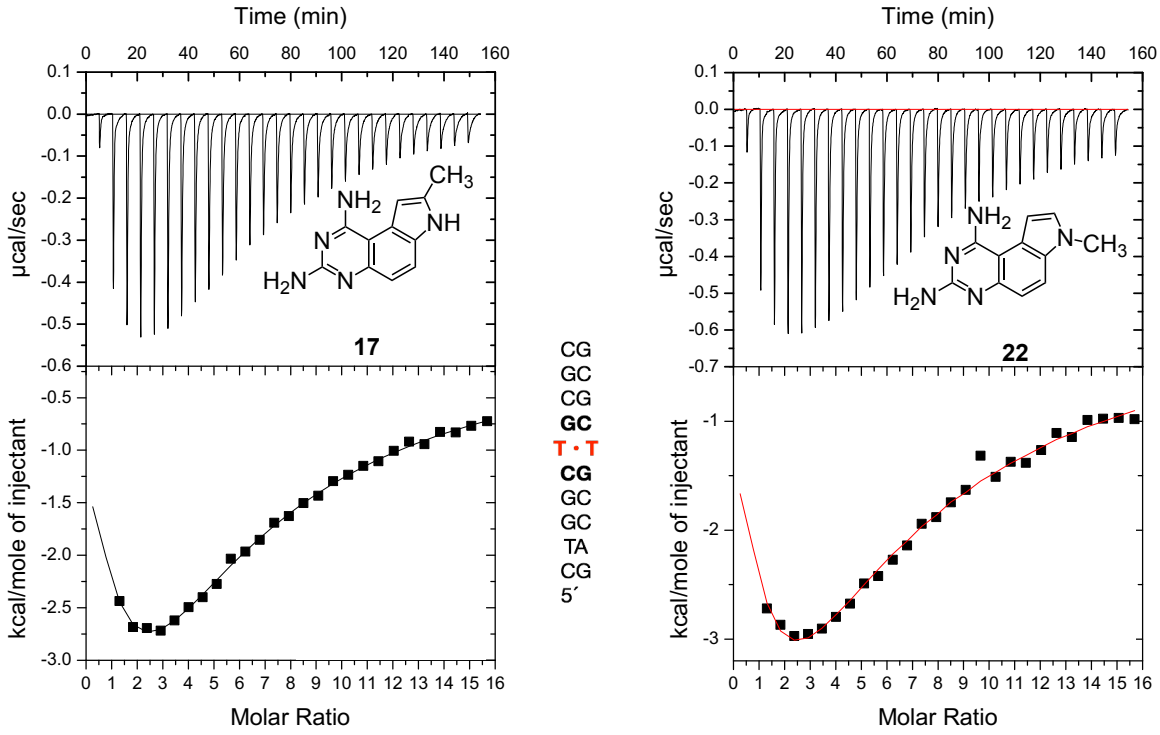
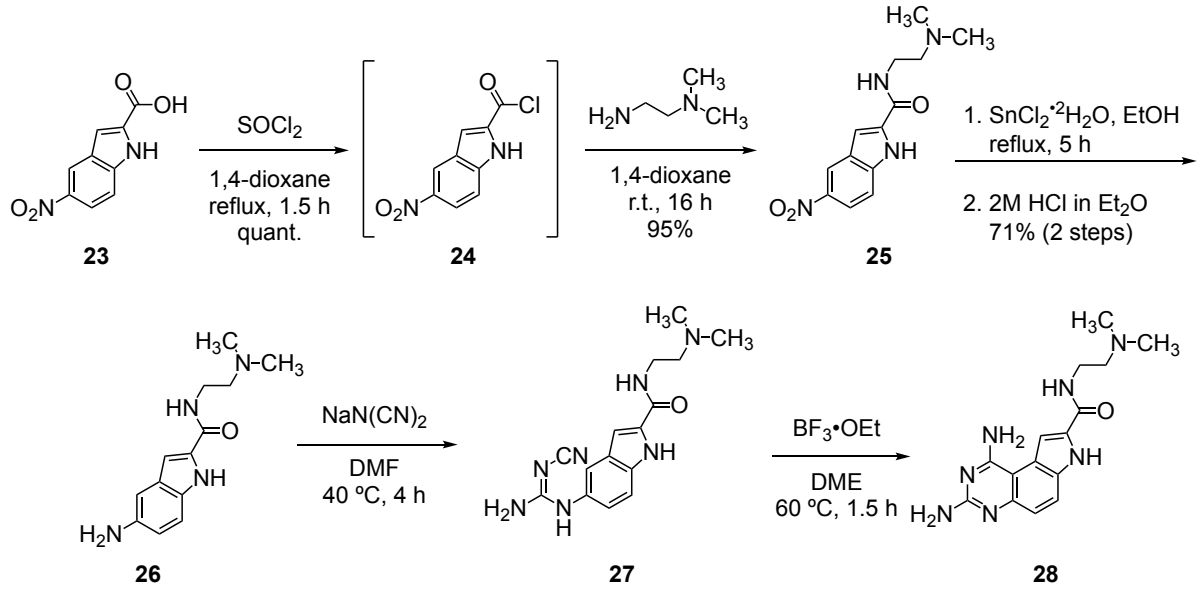
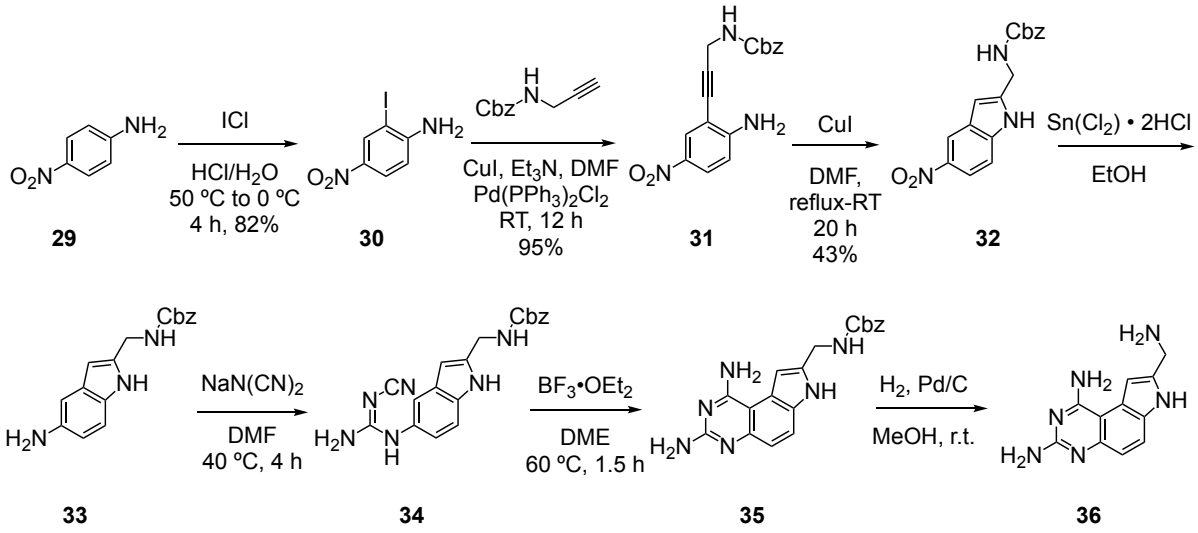


Figure 12. ITC profiles of ligands **17** (left) and **22** (right) with d(CTG)$_2$.



Scheme 3. Synthesis of ligand **28**.

synthesis of ligand **28** as well. This could be useful if a dimeric analog were to be developed, such as one with two pyrroloquinazoline units linked by a polyamine, or incorporating the pyrroloquinazoline unit to another scaffold/backbone. The second proposed ligand **36** shows the incorporation of a simple aminomethyl group. The advantage of such derivative would be the ability to incorporate a broad range of alkyne-bearing groups to the indole core, but its



Scheme 4. Synthesis of ligand **36**.

synthesis involved laborious purifications and was unable to be fully completed. Therefore, ligand **28** is a strong lead derivative for future studies and development.

2.5. Crystallization Efforts of d(CTG)$_2$ –Ligand 11 Complex

In efforts to obtain a definitive and complete insight into the mode of binding of ligand **11** and d(CTG)$_2$, we attempted to grow a crystal for solving the structure of the DNA-ligand complex. Although we were able to obtain some crystals, unfortunately their quality was not suitable for X-ray crystallography. We believe that this could be due to the high percentage (5%) of DMSO used in the crystallization efforts, and not enough DNA and ligand in the samples ([**11**] = 1 mM, [d(CTG)$_2$] = 500 μM). Although the biophysical properties of **11** are promising, we ultimately need to prepare derivatives with better aqueous solubility not only for preparing crystal structures but also for possible future *in vitro* and *in vivo* studies.

2.6. Conclusions and Outlook

We have identified a new lead d(CTG)-binding compound **11** from a fluorescence anisotropy screen performed by Yen-Jun Ho in 2013. Although originally thought to bind r(CUG)^{exp} to inhibit the MBNL1–r(CUG)^{exp} complex, ITC binding studies show that it instead binds d(CTG)$_2$ with low micromolar affinity and possibly in a cooperative manner. We have observed some selectivity toward DNA containing T–T mismatches, but a number of confirmatory studies remain to be done. Most importantly, the exact mode of binding of ligand **11** and its stoichiometry need to be determined, for which the ITC studies contradict the Job analysis. For instance, one can use circular dichroism (CD) titration studies to provide more insight on how the ligand intercalates within the DNA and the stoichiometry. Ultimately, a crystal structure of the d(CTG)$_2$–**11** complex is needed to fully understand the mode of binding.

Nonetheless, in considering ways in which ligand **11** could bind $d(CTG)_2$ cooperatively, a complex that induces base displacement ("flip out") is possible. Indeed, base flipping in normal and mismatched DNA and RNA has been extensively studied and reported,[89-93] with small molecules such as cyclodextrin and metallointercalators also being able to induce base flipping.[25,94,95] It is possible that ligand **11** could induce base flipping due to its larger aromatic surface area, making it less likely to form a base-triplet while being exposed to the extrahelical aqueous environment. Alternatively, two molecules of **11** can bind in one mismatched site, though this is unlikely because the Job analysis shows a 1:1 binding stoichiometry. Finally, if these studies show promise, ligand **11** is also a potential candidate that could be incorporated in the design of fluorescent probes that will help us detect $d(CTG)^{exp}$. So far, only recently has the detection of CAG and CCG repeats have been reported.[96,97]

CHAPTER 3

DEVELOPMENT OF ENFORCED-STACKED INTERCALATORS

3.1. Introduction

3.1.1. Connecting Ligand 3 Binding Mode to Acridine Ligand Design

The acridine-based ligand **JFA** and their analogs that were described in Chapter 1 possessed excellent binding affinity toward oligonucleotides. However, extensive *in vitro* studies[62,73] showed that their cytotoxicity was relatively high; sulforhodamine B (SRB) toxicity studies indicated an IC_{50} = 20.4 ± 0.5 μM for **JFA** in a 24-hour assay.[63,98] In addition, the acridine-based ligands exhibited relatively high nonspecific binding toward dsDNA, as observed in ITC binding experiments with fully-paired dsDNA.[99]

The insights we have gained about the mode of binding of ligand **11** encouraged us to continue our efforts to develop acridine-based DNA-selective small molecules. To better design these ligands, we took into consideration the possibility that we could influence the mode of binding through rational design to favor base flip out. Specifically, designing a ligand that potentially favors the induction of a base flip out could result in better binding affinity and specificity towards $d(CTG)^{exp}$.

3.1.2. Reducing Nonspecific Binding Through Rational Design

We hypothesized that the cytotoxicity results from the off-target binding of unstacked conformers nonspecifically intercalating into the nucleic acids (Figure 13).[63] Thus, we believed that enforcing the acridine and triaminotriazine units to π-stack would result in a further equilibrium shift toward the stacked conformers. To accomplish this, we identified two potential strategies oriented toward improving one of our most promising ligands using rational design: 1) increasing the π-system area of the recognition unit, and 2) employing a macrocyclic

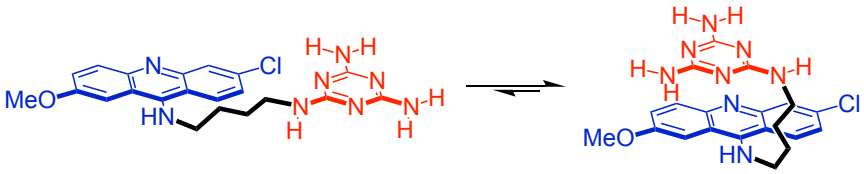


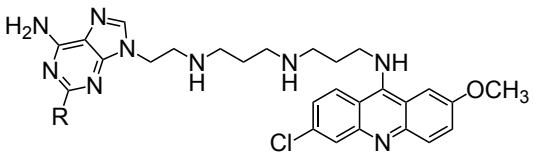
Figure 13. Equilibrium between untacked (left) and stacked conformations of ligand **JFA**.

design. Thus, the overall goal of the work in this chapter was to improve the affinity and selectivity of our acridine-based ligands.

3.2. Increasing the π-System Area of the Recognition Unit

3.2.1. Design and Synthesis of Acridine-Diaminopurine Ligand

Accordingly, we set out to search for new aromatic units that would hydrogen-bond with thymine and were larger than the triaminotriazine units. In 1988, Lhomme and coworkers reported UV/Vis and ^{1}H-NMR studies for a set of acridine-based ligands linked to nucleobases.[100] The intramolecular ring-ring π-stacking interactions between the acridine and adenine or thymine units were found to be very strong in a water/ethanol mixture (95:5 v/v), and in particular the acridine-adenine ligands exhibited almost a two-fold degree of stacking at room temperature, presumably due to its larger aromatic area.[100] It was also observed that the length of the linker reduced the degree of intramolecular stacking. In 1993, Lhomme



37, R = H
38, R = NH_2

Figure 14. Artificial acridine-adenine/2,6-diaminopurine nucleases that cleave apurinic sites in DNA.

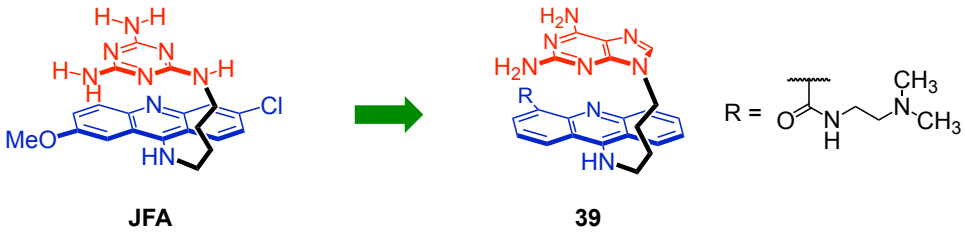
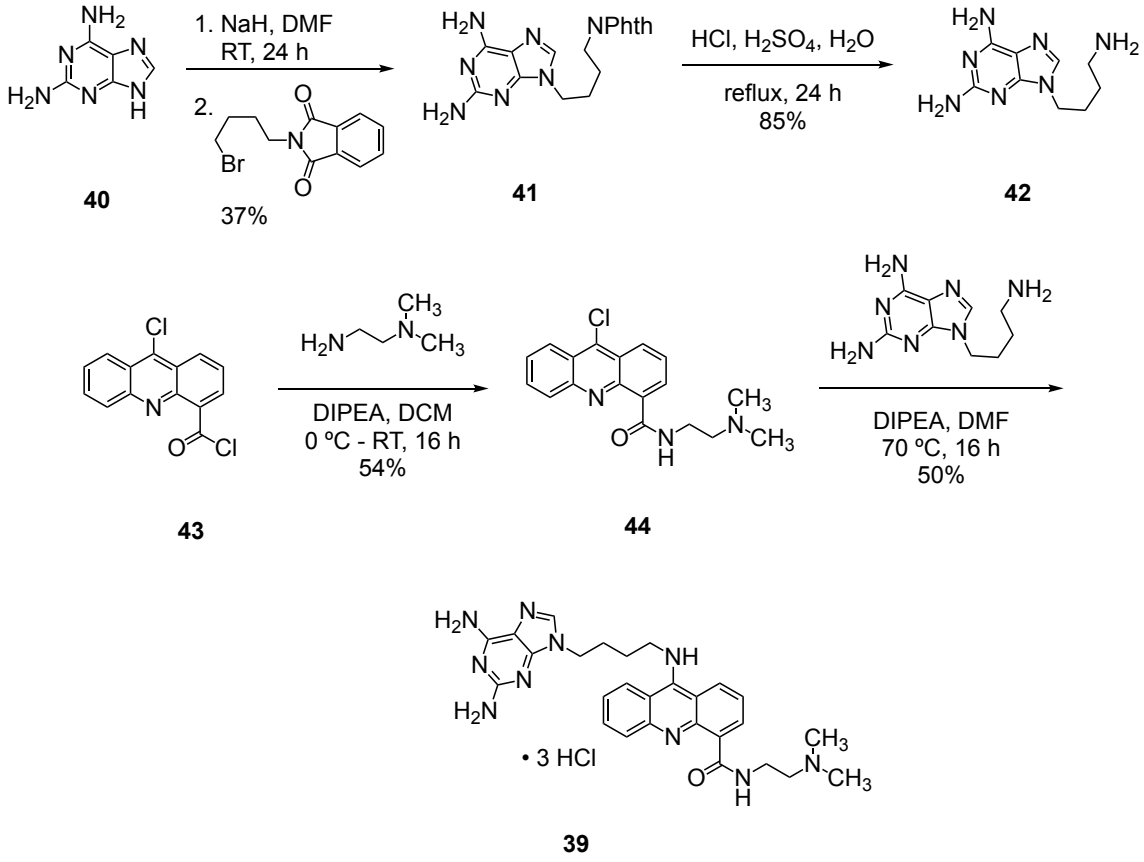


Figure 15. Design progression of ligand **39**.

reported a novel set of acridine-based artificial nucleobases that recognized and cleaved DNA at apurinic sites (Figure 14).[101]

These ligands were also found to exist in the π-stacked conformation at elevated temperatures. Ligand **38** was found to bind DNA 5 times stronger than ligand **37** (K_D = 0.9 μM vs 5 μM), presumably due to 2,6-diaminopurine (DAP) being able to pair with thymine through three hydrogen bonds instead of two as with adenine. Thus, we decided to replace the triaminotriazine recognition unit with DAP (Figure 15). By conserving the 4-methylene linker length, π-stacking between both heteroaromatic units would be heavily favored in water. This would allow us to investigate whether a larger aromatic recognition unit, and hence further favored π-stacked conformation, decreases the nonspecific binding of our acridine ligands to DNA.

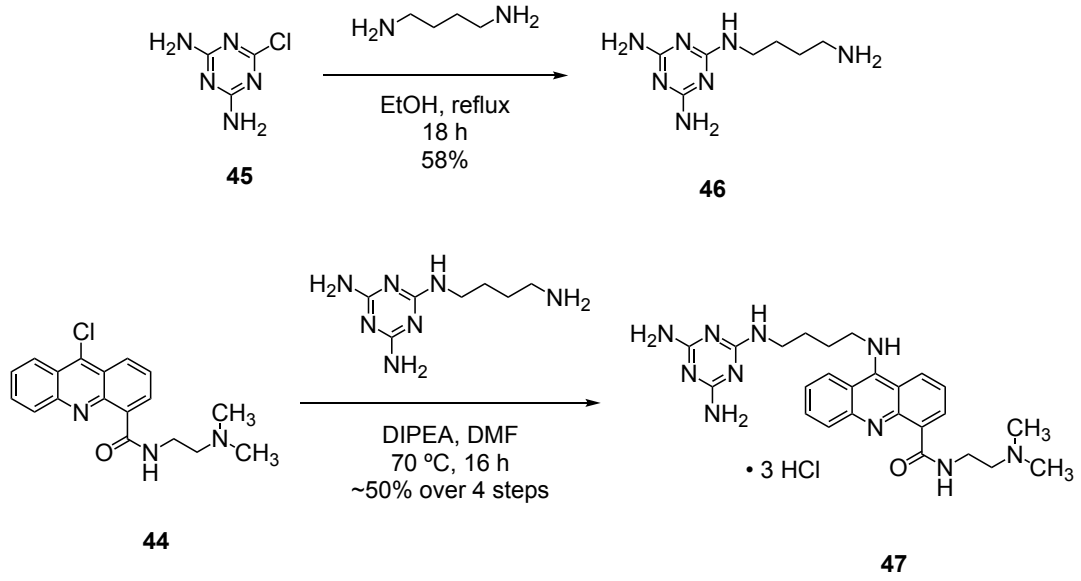
The synthesis of ligand **39** was broadly divided into two parts. The DAP moiety was synthesized based on a procedure for ligand **38**,[101,102] and the water-soluble acridine moiety was synthesized based on our previously reported procedure and introduced at the final stage.[103] As described in Scheme 5**,** the key diaminopurine intermediate **42** was prepared from DAP following the method reported by Leonard for the alkylation of adenine.[104] Introduction of the aminoalkyl chain was achieved by alkylation of the purine with 4-bromobutylphthalimide in DMF the presence of sodium hydride. The reaction took place regioselectively at position 9 (44% yield); no alkylation at the N(7) position was observed.



40 **41** **42** **43** **44** **39**

Scheme 5. Synthesis of DAP-acridine ligand **39**.

Deprotection of the amine was achieved in 85% yield by hydrolysis of the phthalimide group in acidic conditions (HCl/AcOH/H_2O). The purification of intermediate **42** turned out to be a major challenge due to its high polarity. Although we attempted a number of purification methods, ultimately, we had to perform an extraction from water (pH 12) and 10:1 (v/v) chloroform:isopropanol and subsequently use flash column chromatography. The acridine intercalating moiety **44** was prepared *via* an Ullmann-type coupling of anthranilic acid and 2-chlorobenzoic acid,[105] followed by oxidation with thionyl chloride to afford the chloroacridine **43**, and then amidation of the acyl chloride with *N*,*N*′-dimethylethylenediamine in an overall yield of 53%. Finally, the coupling of **42** and **44** was performed in DCM with DIPEA at 65 ºC as previously described[64] to afford **39** in 54% yield. Using similar conditions, the



Scheme 6. Synthesis of control ligand **47**.

alkyltriaminotriazine **46**[62] and acridine **44** were coupled to produce control ligand **47** (Scheme 6).

3.2.2. Biophysical Studies of DAP-Acridine Ligand

To quantitatively determine the affinity and selectivity of ligand **39,** isothermal titration calorimetry (ITC) experiments were performed. Ligand **39** was titrated into solutions of

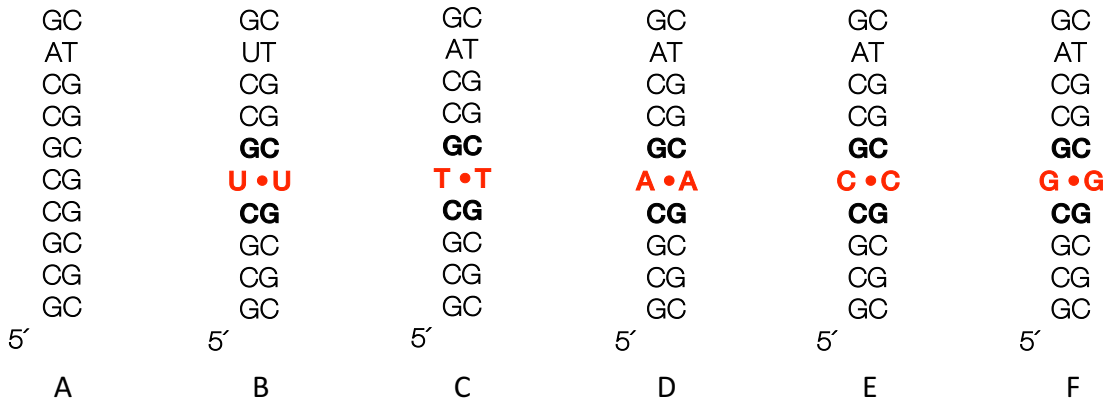


Table 3. List of oligonucleotide duplexes used in biophysical studies.

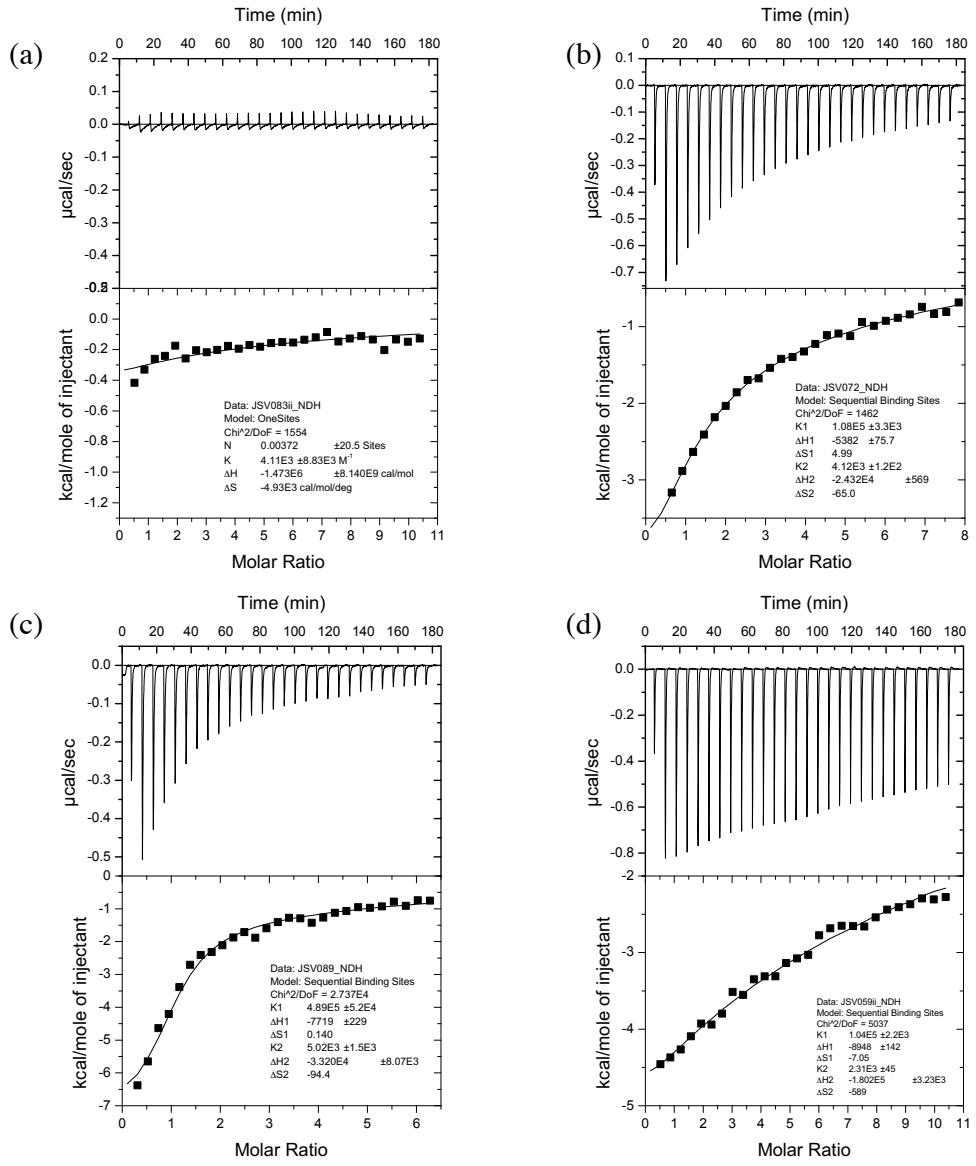


Figure 16. ITC binding isotherms of **39** to r(CUG)$_2$ (a), d(CTG)$_2$ (b), d(CCG)$_2$ (c), and d(CAG)$_2$ (d).

various DNA and RNA 10-mer targets: r(CUG)$_2$, d(CTG)$_2$, d(CAG)$_2$, and d(CCG)$_2$ (Table 3 and Figure 16). The resulting data were fit to a two-site sequential binding site model, with the second binding K_D (200–400 μM) attributed to nonspecific binding.[62] Ligand **39** was found to bind to d(CTG)$_2$ with low micromolar affinity (K_D = 8.9 ± 0.6 μM), but exhibited very weak binding interactions with r(CUG)$_2$ and nonspecific binding (likely nonspecific intercalation) to d(CAG)$_2$ (Figure 16). In addition, the absorption spectrum of **39** was measured at 20 μM

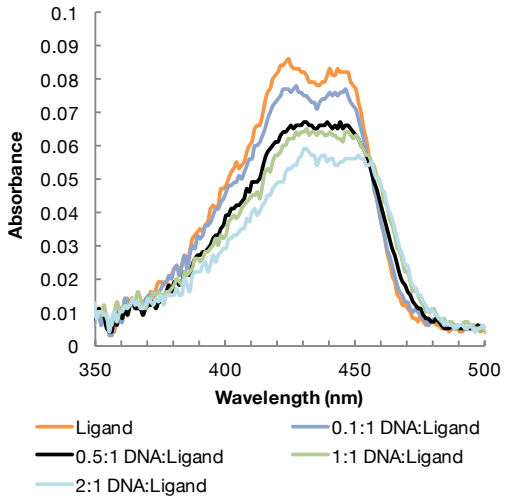


Figure 17. UV-Vis absorption spectrum of **39** and sequence d(CTG)$_2$ at different molar ratios.

with d(CTG)$_2$. It exhibited a 30% decrease in the absorption and a redshift of 10 nm, indicative of intercalation of the ligand into the DNA (Figure 17).[106] However, further ITC experiments revealed a low micromolar affinity ($K_D \sim 3\ \mu M$) to d(CCG)$_2$ (Figure 16c). Although not expected, it was not a complete surprise as the DAP unit can form two hydrogen bonds with cytosine. The similarly strong binding of **39** to d(CCG)$_2$ could suggest that it is selective for pyrimidine mismatches in DNA. However, the C–C mismatch is one of the least stable mismatch pairs in DNA and very weakly hydrogen-bonded.[107] In fact, it has been reported that it can become unstacked from the DNA helix and adopt extrahelical locations.[108] Thus, the weakly mismatched C–C basepairs could facilitate the extrahelical displacement ("flipping") of a mismatched cytosine by the DAP while it π-stacks within the DNA duplex and hydrogen-bonds with the other cytosine (see below and Section 3.2.3).

The ITC isotherms did not show a clear curve inflection point to visualize a binding stoichiometry of the ligand-DNA complex. Thus, using the inherent fluorescence of the acridine moiety of **39**, a Job analysis was carried out at a total concentration of 40 μM, indicating a 1:2 d(CTG)$_2$:**39** binding stoichiometry (Figure 6a). A similar Job plot was obtained when **39** was examined with d(CCG)$_2$. When a Job analysis is performed a lower total concentration of 10 μM, a 1.2:1 stoichiometry was observed. These results suggest that a second weak binding site with a K_D in the high-μM range. Although the second binding

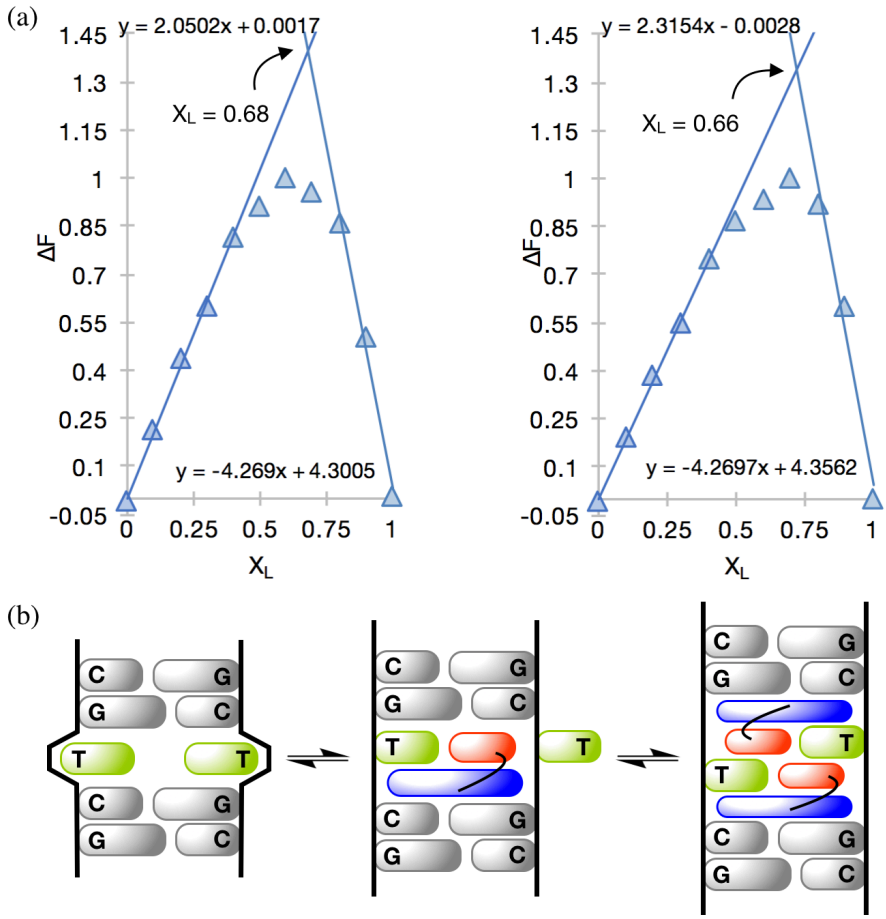


Figure 18. (a) Job plot analysis of **39** with oligonucleotide sequence C (left) and E (right) at a total concentration of 40 μM. The left and right lines in each graph were fitted with a linear model using the first 5 and last 3 data points, respectively. Job analyses were duplicated and were within ± 15%. (b) Cartoon representation of a potential mode of binding with 2 molecules of ligand **39** and one mismatched DNA site (X = T or C) based on Job analysis.

constant obtained from the ITC experiments ($K_{D(2)}$) was originally thought to be associated to nonspecific interactions to the DNA,[109] this binding constant is likely to be from the second binding interaction at the mismatch. These observations add evidence that the DAP unit of ligand **39** potentially induces a base flipping in DNA, but the binding stoichiometry observed of 2:1 needs to be considered. (see Section 3.2.3).

Finally, to assess whether the larger recognition unit had an effect on the ligand's nonspecific binding, ITC experiments were performed using dsDNA and ligand **39** and control **47 (**Figure 7). Unfortunately, a similar titration isotherm was obtained for both ligands,

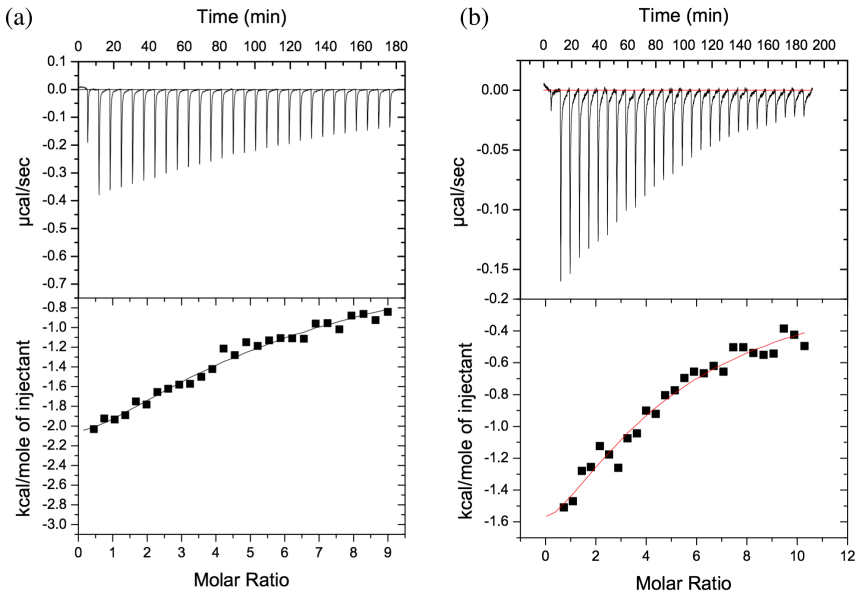


Figure 19. ITC binding isotherms of ligands **41** (a) and **44** (b) to sequence A.

suggesting that the larger aromatic recognition unit does not enforce the intramolecular π-stacking enough to prevent nonspecific intercalation of the ligand to DNA.

3.2.3. Conclusions and Outlook

A new stacked intercalator ligand has been developed with a larger heteroaromatic recognition unit. Low micromolar affinities were in a 2:1 binding stoichiometry were observed for ligand **39** and d(CTG)$_2$. In addition, ligand **39** showed the same stoichiometry and moderate binding affinity for d(CCG)$_2$. Negligible binding was observed with duplex strands containing U–U, A–A, or G–G mismatches. Although not entirely selective of T–T mismatches, the 2,6-diaminopurine unit can be incorporated into future ligands to target C–C mismatches, in addition to single base bulges (C or T). The nonspecific binding of **39** and **47** toward a 10-mer duplex DNA was assessed and compared through ITC studies. Both DAP ligand **39** and control **47** exhibited similar isotherm profiles, suggesting the larger 2,6-diaminopurine recognition unit did not reduce the nonspecific intercalation of the ligand throughout the DNA duplex.

Further studies to need to be conducted to study if ligand **39** induces a base flip out at low concentrations. For instance, regarding the T–T mismatch, oxidation of the C(5)-C(6) double bond of the T base with potassium permanganate would be more feasible for the T base in the extrahelical position compared to that in the intrahelical position of the duplex.[110] Some models of preliminary modes of binding can be envisioned, such as one ligand binding to one mismatched nucleobase and produce a pseudo bulge that allows the binding of another ligand from the opposite direction (Figure 18b), similar to a proposed zigzag intercalation of two naphthyridine rings by Nakatani.[111]Although this mode of intercalation would violate the nearest neighbor exclusion principle,[13] it should not be excluded. Obtaining a crystal structure would provide conclusive information on the mode of binding.

3.3. Employing a Macrocyclic Design

Although the intramolecular ring-ring π-stacking interactions in acridine-purine ligands had been reported to be very strong and exhibit an almost two-fold degree in stacking over acridine-adenine,[100] we did not observe a reduction of nonspecific binding toward DNA (*vide supra*). Next, to further improve the affinity and selectivity of our acridine-based ligands, we proceeded to develop enforced stacked intercalators through our second strategy: a macrocyclic design.

3.3.1. Macrocyclic Bisintercalators

Macrocyclic bisintercalators, also known as cyclobisintercalators, are a unique class of DNA ligands, and one of the most extensively studied macrocycles belonging to the class of heterophanes[112] and cyclophanes.[113] In the late 1980s, Lehn developed macrocyclic bisintercalators consisting of homo- or heterodimeric heteroaromatic units (*e.g.*, acridine, naphthalene, phenanthridine, phenazine, anthracene, imidazolium, bipyridine and/or biphenyl)

connected by two relatively short flexible linkers, and can adopt a semi-closed, intramolecularly π-stacked conformation in the free state.[7] These molecules were initially shown to bind metal ions[114] and, due to their topology, were also able to "sandwich" (*i.e.*, encapsulate) other flat aromatic guests through intercalative supramolecular structures.[115]

In recent decades, several cyclobisintercaland derivatives have been effectively utilized for host-guest and undergo selective with various biomolecules (*e.g.*, nucleotides, proteins, amino acids).[19,116] Zimmerman reported in 1989 the first macrocyclic bisintercalator **48** able to bind a DNA duplex (Figure 20a),[117,118] but it was not until 2003 that Teulade-Fichou reported **49**,[30] the first of this type to recognize mismatch sites and induce base-displacement ("flipping") in DNA (Figure 20b). The two heteroaromatic units of the macrocyclic bisintercalator sandwich a neighboring base pair, with one unit being inserted in an abasic site[119] or flipping a mismatched base pair out of the DNA helix[30] A series of derivatives were developed to investigate their mismatch-binding properties, showing good selectivity for T–T and T-C mismatches (though not between these mismatches).[31] More recently, Nakatani has

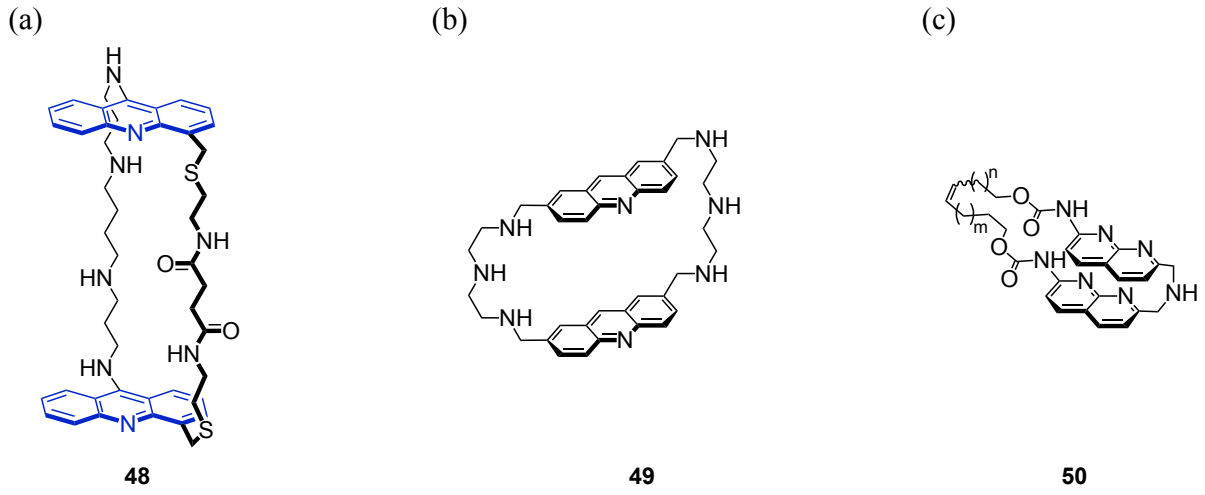


Figure 20. Macrocyclic bisintercalators **48** (a), **49** (b), and **50** (c).

developed a series of macrocyclic bisnaphthyridine dimers **50** that bind DNA bulges[120] and mismatches (Figure 20c).[121]

3.3.2. Macrocyclic Ligand Rational Design

The design of the macrocyclic ligands required the following: 1) an appropriate handle for the attachment of the additional linker, and 2) suitable linker lengths to enforce π-stacking between the two aromatic heterocycles while allowing adjustment to the binding site. As mentioned earlier, our lab had designed ligand **JFA** with an acridine intercalating unit to provide the hydrophobic driving force for binding and a triaminotriazine unit as a Janus-wedge to form a base triplet with U–U or T–T mismatches.[62,73] The carbon linker between the acridine and triaminotriazine units was important in offering the conformational flexibility needed for the concurrent binding of these two units to the target mismatch. Although the two heterocycles are believed to π-stack, there still exists an equilibrium between the stacked and unstacked conformers (Figure 13). Thus, to structurally enforce the π-stacking of the acridine and triaminotriazine units, an additional carbon linker was introduced between the two heterocycles (Figure 21).

The introduction of an additional linker required modifying the acridine with a reactive functional group and the triaminotriazine recognition unit without affecting its hydrogen-bonding functionality in binding the mismatched sites. We also wanted to place linkers on

JFA **A4D3**

Figure 21. Structural progression from acridine-based ligands to macrocyclic bisintercalators.

opposite sides of the acridine unit to favor π-stacking with the triaminotriazine unit as much as possible. We had previously reported more water-soluble derivatives of ligand **JFA** by using a different intercalator, *N*-[2-(dimethylamino)ethyl]acridine-4-carboxamide (DACA), which gave the option to use a carboxylic acid group at the 4-position that could be amidated with an amine linker.[65,103] It had been previously reported that placing a carboxylic acid group in the 3-position is synthetically challenging because it produces an inseparable mixture with the isomer at the 1-position.[122] In addition, using the methyl scanning method we had showed that the amino groups on the triaminotriazine unit of **JFA** could bear an alkyl substituent without

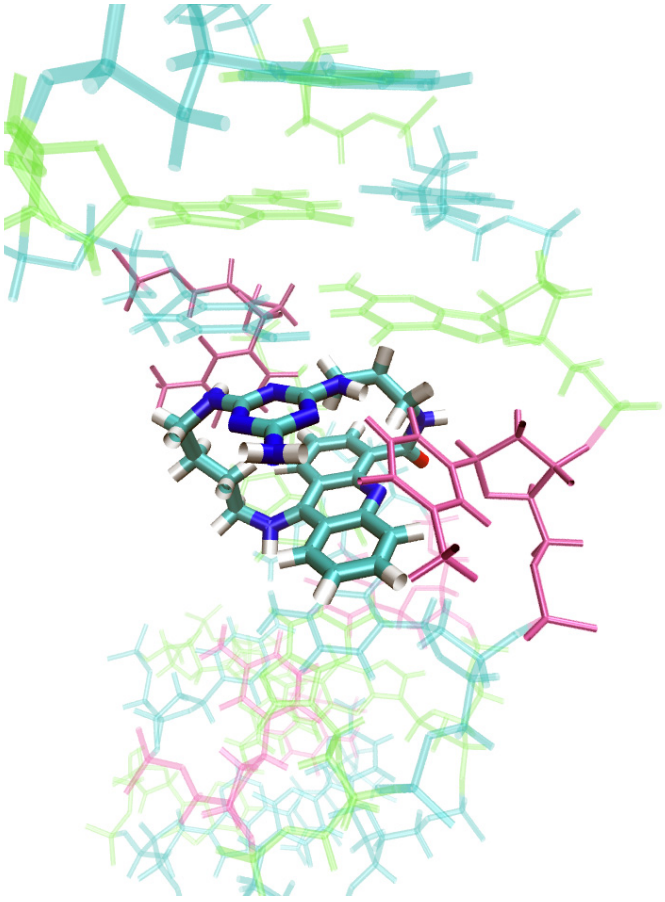


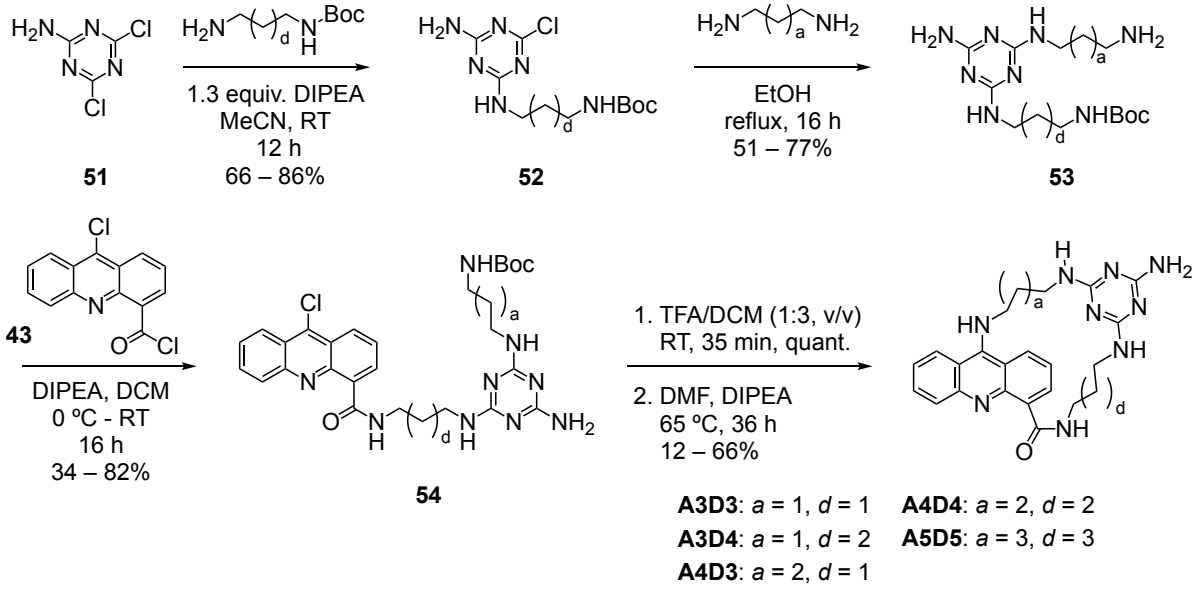
Figure 22. Energy-minimized complex modeling showing ligand **A4D3** binding to a mismatched CTG site. The triaminotriazine unit of the ligand hydrogen-bonds to one of the mismatched thymines. (T = magenta, C = teal, G = light green.) Computational modeling performed by Lauren Hagler.

affecting its binding toward $d(CTG)_2$.[73] Therefore, we decided to introduce the additional linker connected at one of the triaminotriazine amines and at the 4-position of the acridine unit.

Whereas most macrocyclic bisintercalators recognize nucleic bases and mismatched sites by stacking between two separate base pairs, the macrocycles in this work were designed with short linkers to form a macrocycle small enough so that intercalation is only possible between the GC base pair and the mismatched base pair site and the triaminotriazine unit either forms a base triplet or cause a base flip out. In order to assess the length of the linkers and π-stacking between the acridine and triaminotriazine units, Lauren Hagler, a graduate student in the Zimmerman Group, performed molecular modeling computations. The macrocyclic conformation in the presence of water was modeled in the computational modeling program Molecular Operational Environment (MOE). The DNA sequence synthesizer in MOE was used to create the DNA duplex containing CTG mismatched sites. A model macrocycle with 3- and 4-methylene linkers (**A4D3**) to provide enough conformational flexibility was built and energy-minimized using the build function of the software. A final energy minimization of the entire **A4D3**–DNA complex produced the structure in Figure 22. This energy minimized structure of the ligand shows its enforced stacked conformation, where the triaminotriazine and acridine units are π-stacked. As can be seen, one side of the recognition unit of the ligand interacts with one of the mismatched T-bases (shown in magenta), while the other T-base is pushed out into an almost a flipped-out position. This helped us choose linkers that would provide enough flexibility to allow the triaminotriazine unit to adjust and bind at the mismatched bases. Likewise, **A4D3** was modeled with a crystal structure for an RNA sequence containing CUG mismatched sites (PDB ID: 3gm7), and no low-energy minima was observed, suggesting no global minima and possibly no binding of **A4D3** to the RNA.

3.3.3. Synthesis of the Macrocyclic Ligands

A series of five cyclobisintercalators were synthesized with different linker lengths, ranging from five to eight atoms (Scheme 7). The goal of this small library was to determine the optimal linker lengths separating the two heteroaromatic units, and to study whether the extra flexibility affects binding selectivity to mismatched bases. The synthesis of the ligands was divided in two parts: synthesizing the recognition moiety with both linkers attached, followed by coupling to the acridine intercalating unit. Attaching the linkers to the acridine unit first was also considered, but due to the slow hydrolysis of 9-aminoacridines at the N-9 position, it was best to limit the number of steps performed with this unit in place. By controlling the temperature, 2,4,6-trisubstituted triazines can be prepared by sequential selective addition of nucleophiles (*e.g.*, amines).[123] Thus, triazine **51** was prepared by treating 2,4,6-trichloro-1,3,5-triazine[124] with ammonia and water at 0 ºC. Addition of the appropriate mono Boc-protected diamine at room temperature afforded triazine **52**. Finally, addition of a

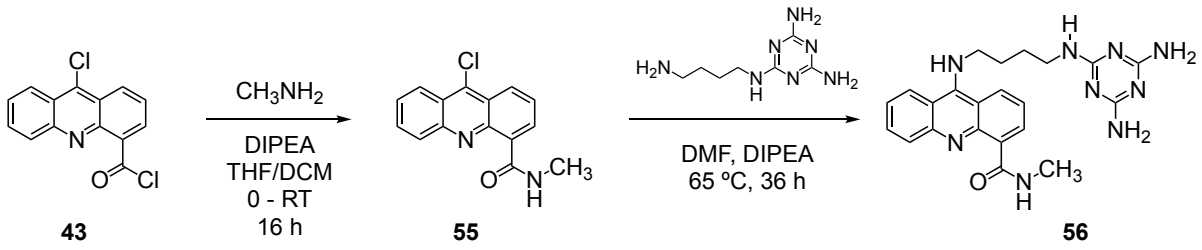


Scheme 7. Overview of the synthetic procedures for the macrocyclic ligands.

solution of **52** to a large excess of the appropriate diamine afforded the triazine recognition moiety **53**. Its purification, and in particular the removal of the excess diamine, was challenging and ultimately, silica flash chromatography was used with DCM:MeOH:NH_4OH (10:1:0.1 v/v) as the eluent. The free amine was reacted with the acyl chloride group in acridine **43** to produce intermediate **54**. After purification and removal of most impurities, intermediate **54** was briefly treated with TFA to remove the Boc group. Finally, intramolecular cyclization was performed under very dilute conditions (~1 mM), affording the corresponding macrocycles in fair yields. The cyclization conditions were especially challenging, as solvents and reagents needed to be extremely anhydrous due to the potential of any water interfering with the reaction at the 9-chloro position of the acridine. Using similar conditions and previously reported conditions[62], the noncyclic control ligand **56** was prepared (Scheme 8).

3.3.4. Biophysical Studies of the Macrocyclic Ligands

The binding affinity and selectivity for T–T, U–U, A–A, G–G, C–C mismatches, and dsDNA were studied using ITC. The macrocycles were titrated into solutions of various DNA/RNA 10-mer duplexes containing a single mismatched CXG site (X = A, T, C, G, or U) were used (Table 4). The binding isotherm data was fit to a one-site binding model. All the macrocycles showed similar binding affinity to d(CTG)$_2$ (K_D = 19–35 μM) and apparent



Scheme 8. Synthesis of noncyclic acridine control **56**.

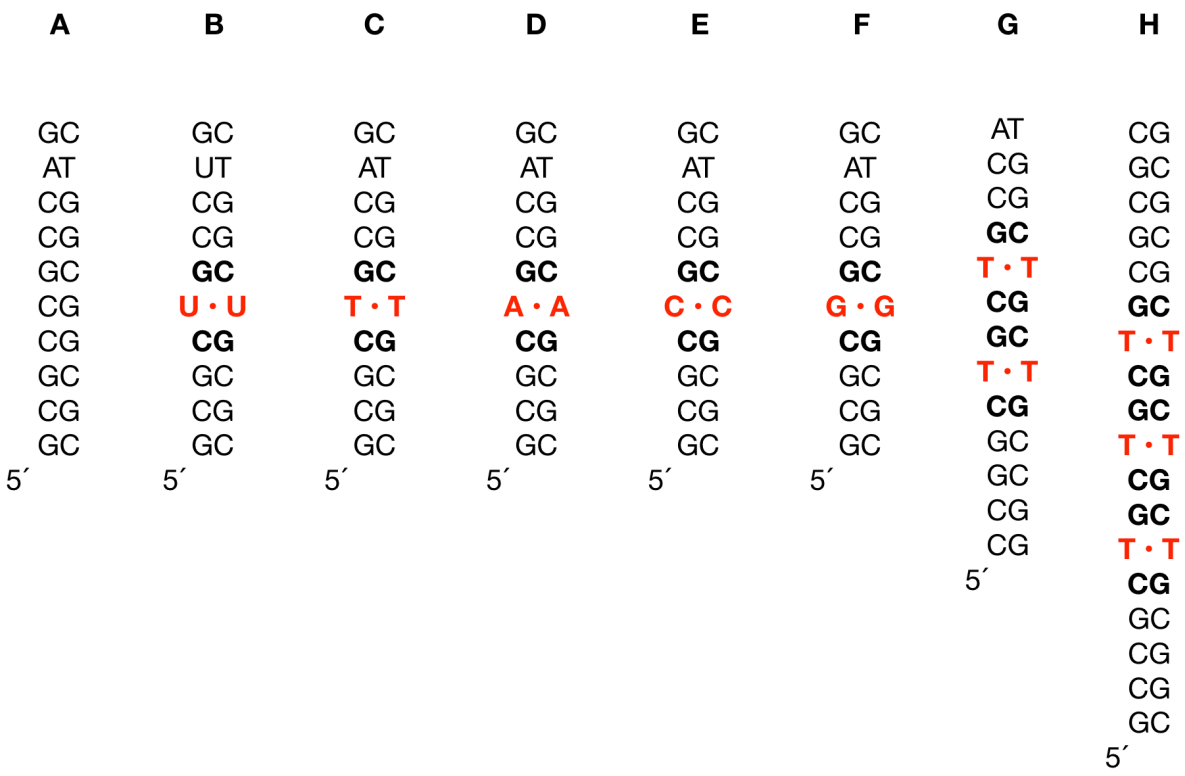


Table 4. Oligonucleotide duplexes used for biophysical studies. The structures shown represent the most stable structures predicted by m-fold.

stoichiometries of 1:1, whereas none of the macrocycles showed detectable binding toward U–U, G–G, and A–A mismatches (Table 5). Control ligand **56** bound moderately to U–U mismatches, slightly weaker than the reported value for **JFA** (~2 μM, although this was calculated using a two-site sequential model). We were particularly interested in the nonspecific binding toward duplex DNA of the macrocycles versus the noncyclic analog **56**. Control ligand **56** bound dsDNA nonspecifically with an affinity of 69 ± 25 μM. On the other hand, none of the macrocycles showed detectable binding affinities toward duplex DNA (Figure 11). Due to the instrument's upper detection limit, it is estimated that the binding is >1 mM.

In addition, only three macrocycles, **A3D4, A4D4,** and **A5D5**, showed weak binding toward C–C mismatches (sequence E; 404 μM, 155 μM, and 125 μM respectively), similarly to control **56** (159 ± 22 μM) and about 20-80 times weaker than **JFA** (5.0 ± 0.2 μM).[62] As observed for acridine-DAP ligand **3**, because the C–C mismatch is very unstable,[107] the macrocycles **A3D4** and **A4D4** could be displacing a mismatched cytosine into an extrahelical position. Overall, although the T–T mismatch binding-affinity was slightly reduced, these results show that the nonspecific binding toward dsDNA was greatly reduced with a macrocyclic structure.

To further analyze and quantify the reduction in nonspecific binding between the macrocycles and control **56** toward dsDNA, reverse binding titrations were performed using fluorescence spectroscopy. The fluorescence change due to binding was monitored as 1-μL aliquots of DNA at various concentrations were manually titrated into solutions of [**A4D3**]$_0$ = 0.2 μM, and the data was fit to a one-site binding model using the software Prism. A clear decrease was observed when the d(CTG)$_2$ was titrated into solutions of **A4D3** and **56**, with K_D values of 0.8 ± 0.6 μM and 3.5 ± 2 μM, respectively. The binding affinities of the control

Duplex	Target	**A3D3**	**A3D4**	**A4D3**	**A4D4**	**A5D5**	**49**
A	Duplex	nb	nb	nb	nb	nb	69 ± 25
B	d(CTG)$_2$	19.5	34.9 ± 17	19.8 ± 13	19.2		5.8
C	d(CAG)$_2$	nb	nb	nb	nb	nb	—
D	d(CCG)$_2$	nb	404	nb	155	125	159 ± 22
E	r(CUG)$_2$	nb	nb	nb	nb	nb	17.9

*Average apparent K_D were determined from 2-4 independent experiments, fit using a one-site model. In cases where only two repeat experiments were made, no standard deviation is listed. Nb indicates no detectable binding.

Table 5. Equilibrium dissociation constants* (apparent K_D, μM) of **A3D3**, **A3D4**, **A4D3**, **A4D4**, and **A4D5** to various oligonucleotides determined by ITC.

ligand is comparable to that obtained from ITC; however, the binding affinity for **A4D3** is about 20-fold lower. The reason for this inconsistency is not known at this time.

Furthermore, when titrating solutions of dsDNA into the ligands' solutions, the data obtained was irregular from trial to trial. Given the uncertainty and error involved, no fit could be confidently used to assess the strength of binding of the macrocycles to dsDNA. As an alternative attempt to quantify the difference in nonspecific binding between the macrocycles and the noncyclic control ligand toward dsDNA, ethidium displacement experiments were performed. When macrocycle **A3D3** was titrated to duplex **A** complexed with ethidium, minimal ethidium displacement occurred. However, when the control ligand **56** was titrated to

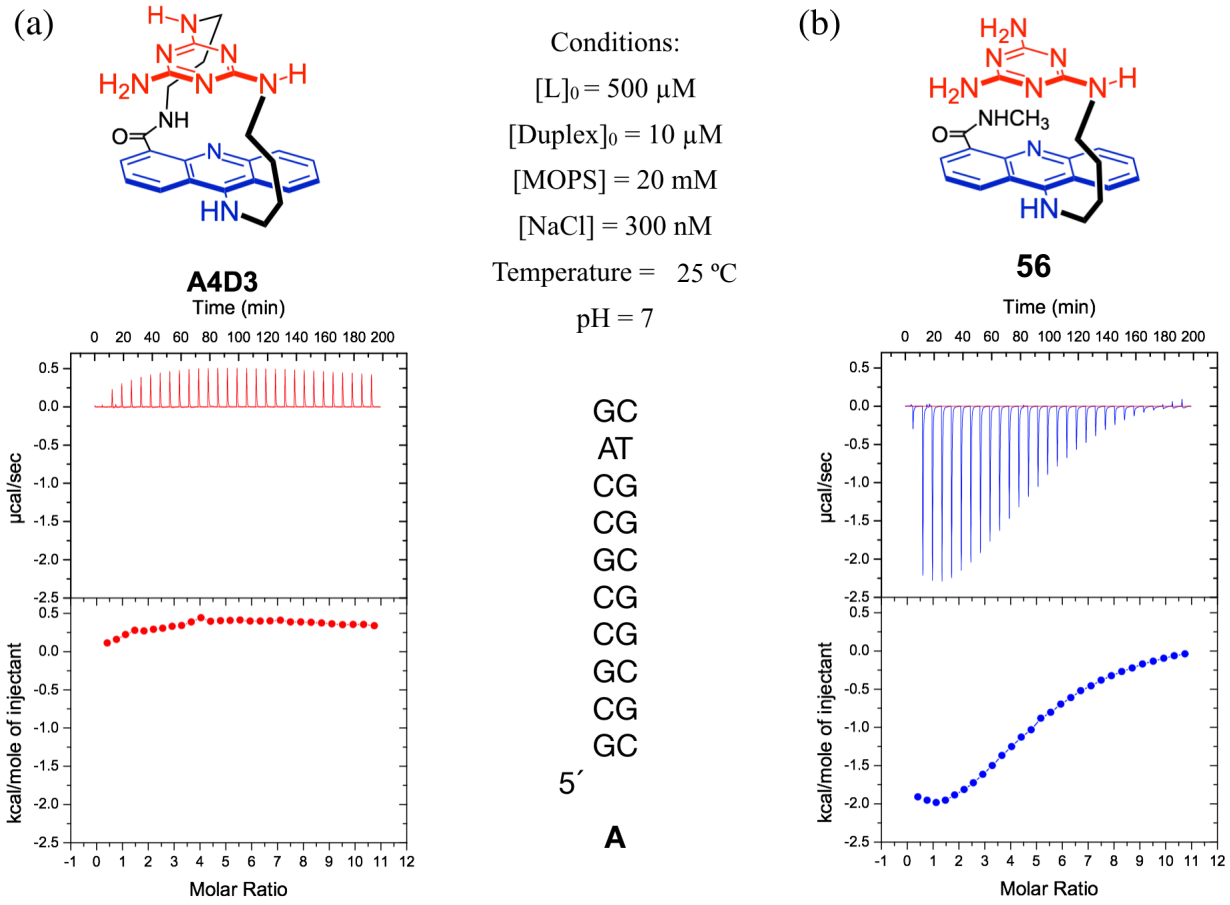


Figure 23. ITC binding isotherms of **A4D3** (a) and **56** (b) to sequence A. The binding isotherm for **A4D3** is representative of the other macrocycles and sequence A (see Appendix).

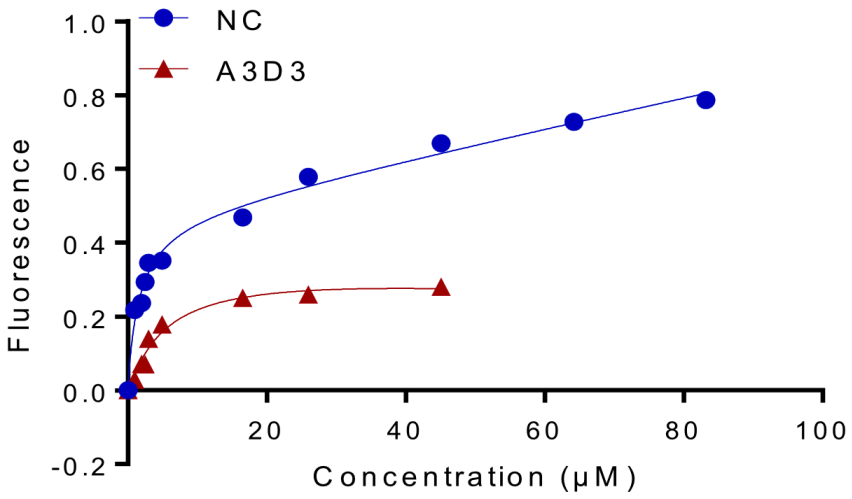
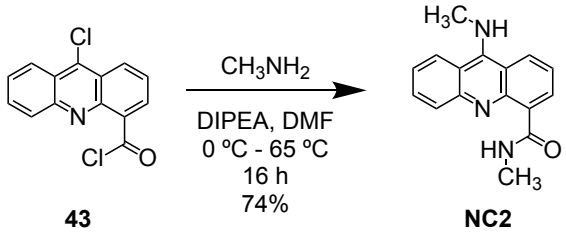


Figure 24. Ethidium displacement experiment in which **A3D3** (red) and **56** (blue) are titrated to ethidium:dsDNA complex, showing the noncyclic control was able to displace ethidium about 4 times more than the macrocycle ligand. Fluorescence experiments were duplicated and the results were within ±5%. Noncyclic control ligand **56** is **NC**.

a ethidium:**A** complex, ethidium was readily displaced (Figure 25). Fitting the data to a one-site model provides an apparent K_D of 1.6 ± 0.5 and 5.8 ± 2 μM for the control and macrocycle, respectively, suggesting the macrocycle can displace ethidium 4 times less than the noncyclic control ligand. This result indicates and confirms the weaker nonspecific binding of **A3D3** relative to the noncyclic control A similar result was obtained from a single trial displacement experiment with **A4D3**.



43 **NC2**

Scheme 9. Synthesis of control **NC2.**

As all the macrocycles had comparatively moderate binding affinities and similar stoichiometries by ITC, we chose **A4D3** as a model ligand and starting point to study the mode of binding toward DNA. Evidence of an intercalative binding with a 1:1 **A4D3**:DNA ratio was obtained in a few ways. First, a Job plot analysis at a total concentration of 50 μM using sequence C gave a maximum concentration of complex with a ligand mole fraction of 0.49, which indicates a 1:1 stoichiometry (Figure 24a) and validates one high-affinity binding site observed in ITC. Also, UV/Vis titration experiments showed a red shift of 5 nm of the visible absorption band of the macrocycle as the DNA:**A4D3** ratio increased from 0.1 to 2, suggesting

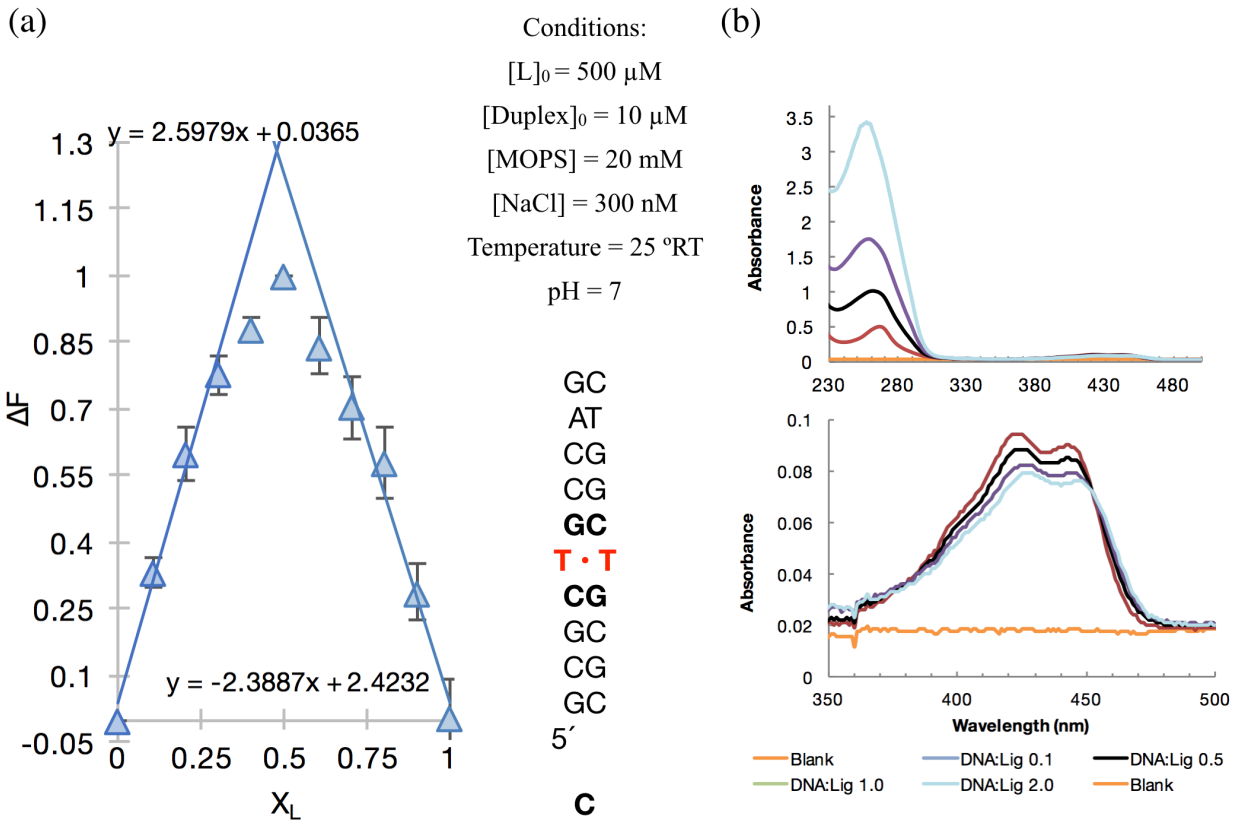


Figure 25. (a) Job analysis of ligand **A4D3** to d(CTG)$_2$ duplex A at a total concentration of 50 μM. The left and right lines in each graph were fitted with a linear model using the first 4 and last 4 data points, respectively, intersecting at a ligand mole fraction of 0.49, indicating a 1:1 stoichiometry. Error bars represent standard deviation of three independent experiments. (b) Absorption spectrum of **A4D3** with duplex C at various ratios. The bottom spectrum is a zoomed-in window from 350 nm–500 nm.

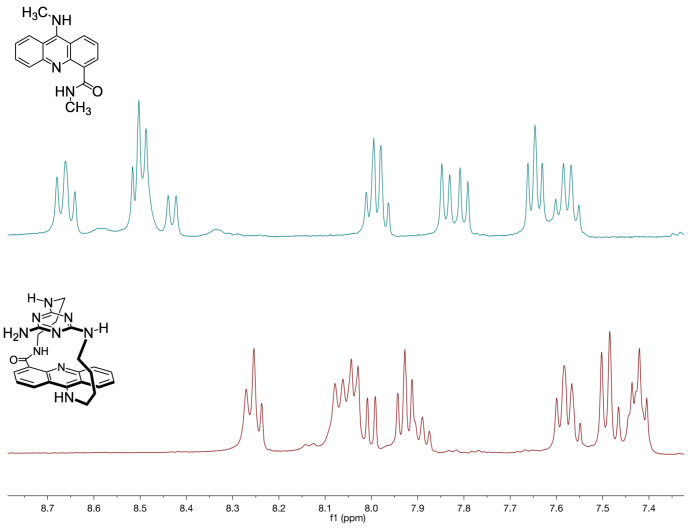


Figure 26. ^{1}H NMR spectra of **A4D3** (green, top) and **NC2** (red, bottom) in D_2O at ca. 200 μM. The observed upfield chemical shifts of **A4D3** indicate intramolecular π-stacking in water.

the ligand was intercalating in the DNA (Figure 24b). (All other macrocycles in the presence

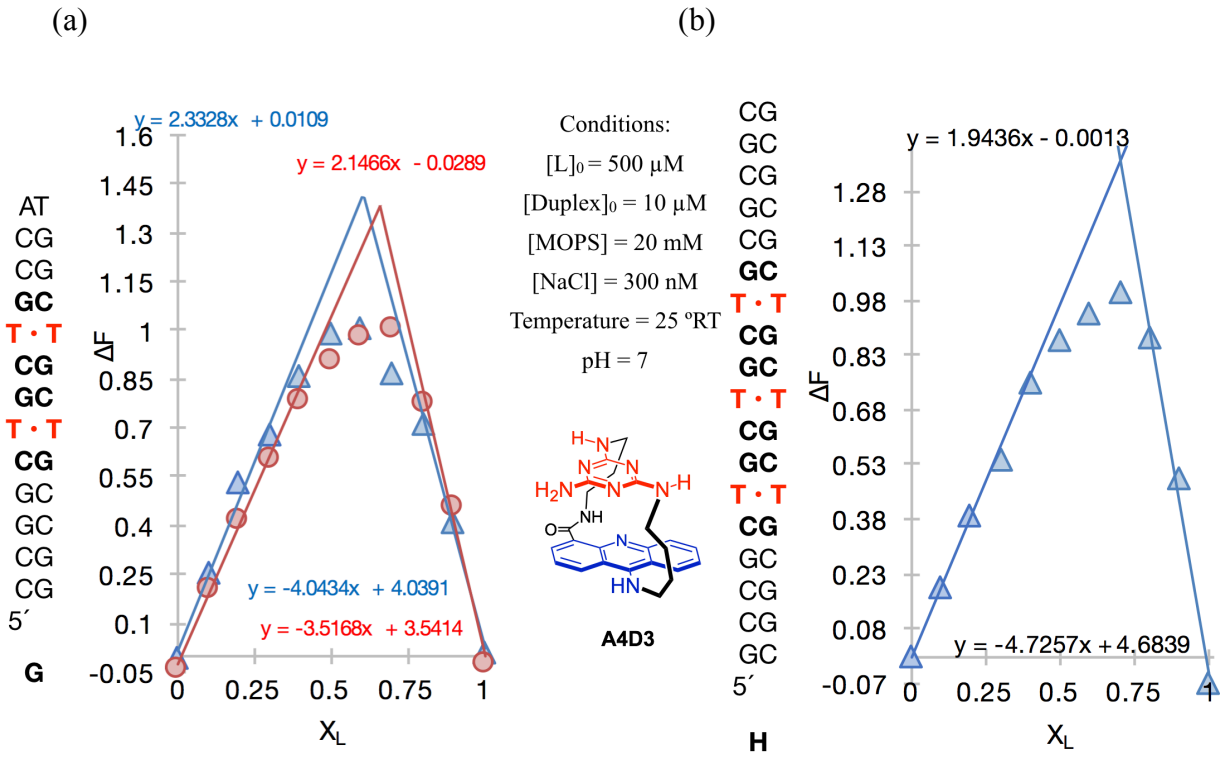


Figure 27. (a) Job analysis of **A4D3** with d(CTG)$_4$ duplex G at a total concentration of 60 μM (blue) and 200 μM (red) indicating a 1.5 : 1 and 1.9 : 1 stoichiometry, respectively. (b) Job analysis of **A4D3** with d(CTG)$_6$ duplex H at 60 μM (right) indicating a 2.1 : 1 stoichiometry. The left and right lines in each graph were fitted with a linear model using the first 4 and last 4 data points, respectively. Job analyses were duplicated and were within ± 15%.

of $d(CTG)_2$ exhibited similar spectra.)

Sequences G and H, containing 2 and 3 CTG sites, respectively, were chosen to examine the ability of **A4D3** to bind multiple sites within a CTG repeat structure. Job analyses at a total concentration of 60 μM to study the binding stoichiometry of **A4D3** to DNA duplexes G and H resulted in binding ratios of 1.5:1 and 2:1 **A4D3**:DNA, respectively (Figure 25a and 25b). It is worthwhile to note that these results are consistent to the reported stoichiometries observed with **JFA**.[62] The proximity in binding sites limits the complexation of two ligands in sequence G. In sequence H, a 2:1 stoichiometry is observed, presumably from two ligands binding at the terminal CTG sites and leaving the central CTG site empty. Indeed, when a Job analysis was performed with **A4D3** and sequence G at a higher total concentration of 200 μM, the binding stoichiometry observed was 1.9:1 (Figure 25a, red). Overall, it appears that the enforced stacking prevents the macrocycles from intercalating nonspecifically because it is not possible for the two adjacent base pairs to separate by the thickness of the two stacked aromatic heterocycles (6.8 Å)[12] except at sites containing a mismatch. All five macrocycles appear to bind to CTG sites with a similar affinity, suggesting the change in freedom of the recognition unit by varying linker lengths does not significantly affect the binding affinity.

Lastly, the intramolecular π-stacking of **A4D3** was studied by ^{1}H NMR in deuterium oxide at low concentrations (ca. 200 μM). As a control ligand, acridine **NC2** was synthesized relatively straightforward (Scheme 4). The acridine protons showed an upfield signal shift for **A4D3** due to shielding, compared to the chemical shift of the acridine protons in **NC2**. This observation indicated intramolecular π-stacking between the two aromatic units in aqueous solution (Figure 26).

3.3.5. Cytotoxicity and Transcription Inhibition Assays with Lead Macrocyclic Ligands

As mentioned earlier, we hypothesized that the cytotoxicity of our acridine-triaminotriazine ligands results from the off-target binding of unstacked conformers nonspecifically intercalating into the nucleic acids (Figure 1). Thus, we wanted to study the effect of a macrocyclic structure on cell viability. The following cell uptake, cytotoxicity and transcription inhibition assays were carried out by JuYeon Lee, a graduate student in the Zimmerman Group. Live HeLa cells were incubated with 30 μM of **A4D3** for 12 hours. Confocal microscopy images provided preliminary semiqualitative evidence that the macrocyclic ligand could pass through the cell membrane and enter the nucleus (Figure 27).

Next, to assess *in vitro* cytotoxicity, sulforhodamine B (SRB) colorimetric assays, an effective and one of the most widely used methods were performed using HeLa cells.[125,126] SRB is a bright-pink aminoxanthene dye known to bind protein components of cells. The amount of SRB bound to proteins is directly proportional to total protein content, thus providing information about total cell mass (*i.e.*, growth). HeLa cells were treated with serial concentrations of 100 μM to 0.2 μM of the five macrocyclic ligands and control **56**. After being incubated at 37 ^{o}C for 72 h, the cells were fixed and treated with SRB. The bound SRB

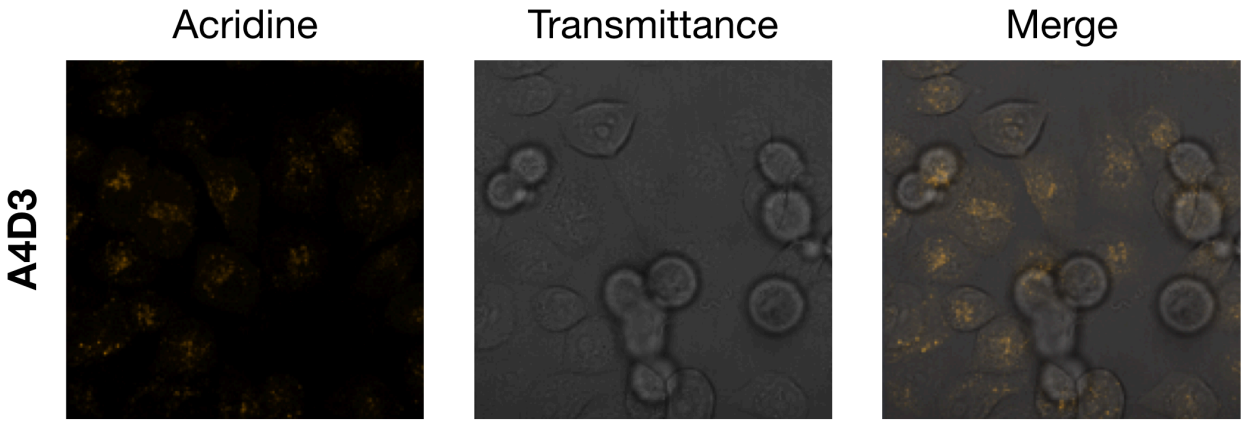


Figure 28. Cellular uptake of ligand **A4D3** into HeLa cells monitored by confocal microscopy. Images of live cells were taken after incubating for 12 h with a ligand concentration of 30 μM. The acridine units of the macrocyclic ligand showed inherent fluorescence (left).

was extracted and its fluorescence was measured. The percentage of cell death was calculated by comparing with untreated samples, and it was plotted against concentrations of the macrocycles. The IC_{50} values, which is the concentration resulting in 50% cell death, ranged from 100 – 12 μM (Figure 28). The control ligand **56** showed a similar IC_{50} value to our initial acridine ligand **JFA** (~20 μM after 24 h).[98] In addition, no observable cell death was found for ligands **A3D3** and **A3D4** in this cell line at concentrations as high as 25 μM for 72 h. However, significant cytotoxicity was observed for **A4D4**, **A4D3**, and **A5D4** at this concentration (Figure 28). This suggests that the cytotoxicity of the acridine-based ligands does not necessarily depend on macrocyclic or noncyclic structure. The only two macrocycles that showed a lower toxicity profile than **56** were **A3D3** and **A3D4**. Ligands **A4D4** and **A4D3**, which contain the same linker length between the acridine and triaminotriazine units as **JFA**, showed a very similar toxicity profile to control **56**.

These results do not support our hypothesis that the cytotoxicity observed in our noncyclic acridine-based ligands is due to their nonspecific intercalation into the nucleic acids. Although the reason for the macrocycles' toxicity is currently not known, there appears to be a trend in the linker lengths. Ligands **A3D4** and **A3D3** with 3-methylene linkers off of the 9-position on the acridine unit exhibit the lowest cytotoxicity. However, **A4D3**, **A4D4**, and **A5D5** with 4-methylene and 5-methylene linkers at the same position exhibit ~5-fold higher cytotoxicity, which is similar to that observed by the noncyclic control **56**. One could argue that the larger size of the molecules, and thus their overall hydrophobicity, has an effect as those with longer

linker lengths exhibit higher toxicity. However, macrocyclic ligands **A3D4** and **A4D3** show every different IC_{50} values (~100 and ~20 μM, respectively).

Although preliminary molecular modeling does not show a clear difference in the way **A3D4** and **A4D3** interact with $d(CTG)_2$, as they both appear to π-stack and hydrogen-bond similarly to the mismatched site, our results suggest that the shorter A-linker (*i.e.*, **A3D3** and **A3D4**) could be affecting the ligands' conformations and toxicity. Further computational and biological studies need to be conducted to investigate the nature of our noncyclic ligands. One possibility is that the noncyclic ligands do not intercalate nonspecifically to a significant extent

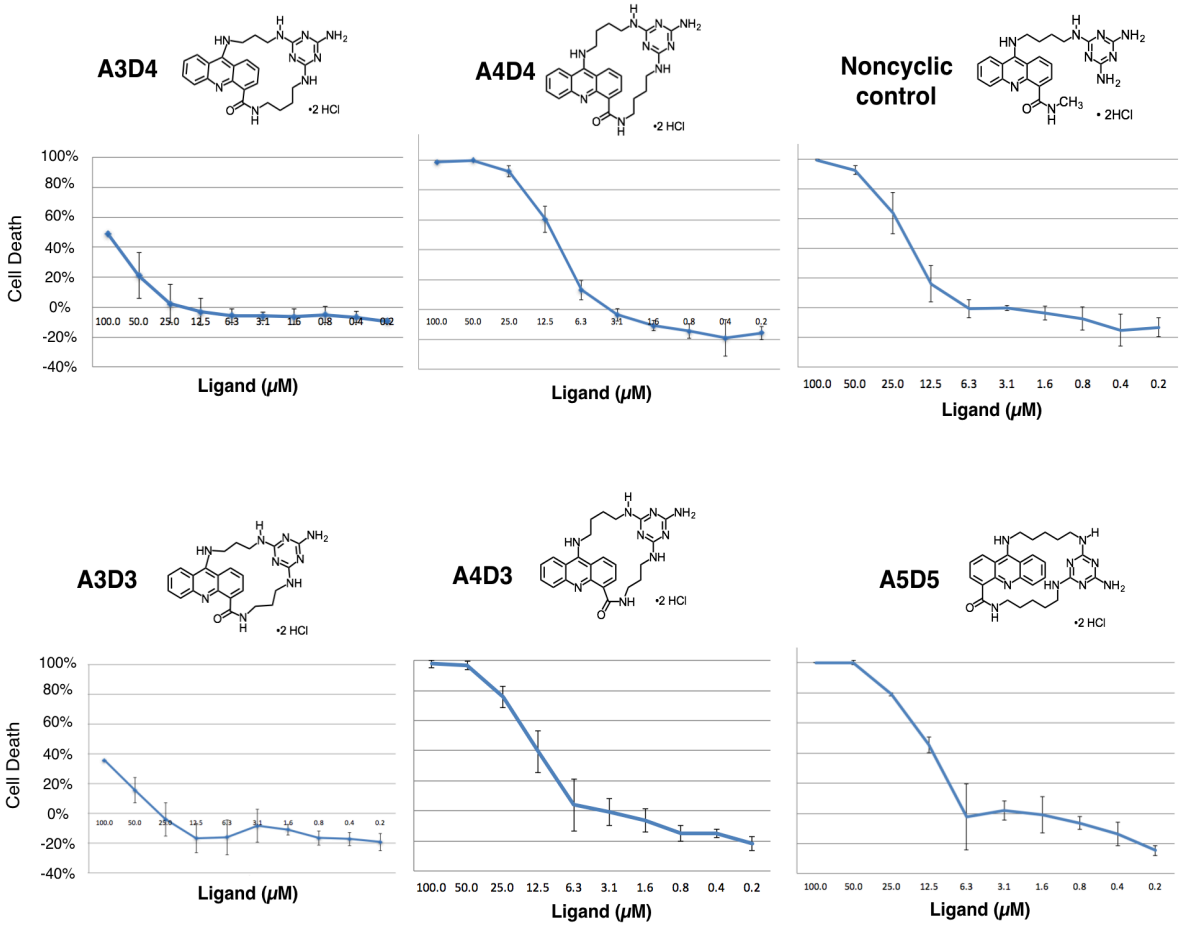


Figure 29. Evaluation of the toxicity of the macrocycles to HeLa cells. The ligands were incubated with cells for 72 h and the cell death was assessed using sulforhodamine B (SRB). Error bars represent standard deviations of three independent experiments.

with dsDNA to cause toxicity at the studied concentrations, but instead interact and interfere with processes in cells that lead to cell death. For example, the macrocycles could be interacting with single-stranded DNA or interfering with proteins involved in DNA repair processes, eventually triggering cell death.

Next, due to the ability of the macrocyclic ligands to selectively bind T–T mismatches, we wanted to assess their potential for (CTG•CAG) transcription inhibition. The *in vitro* transcription inhibition assay used linearized CTG_{74} plasmid and a T7 promoter located in the upstream region of the repeats. Macrocycles **A3D3**, **A3D4**, and **A4D3** were selected based on their binding affinities toward CTG sites and their cytotoxicity profiles. Macrocycles **A3D3**,

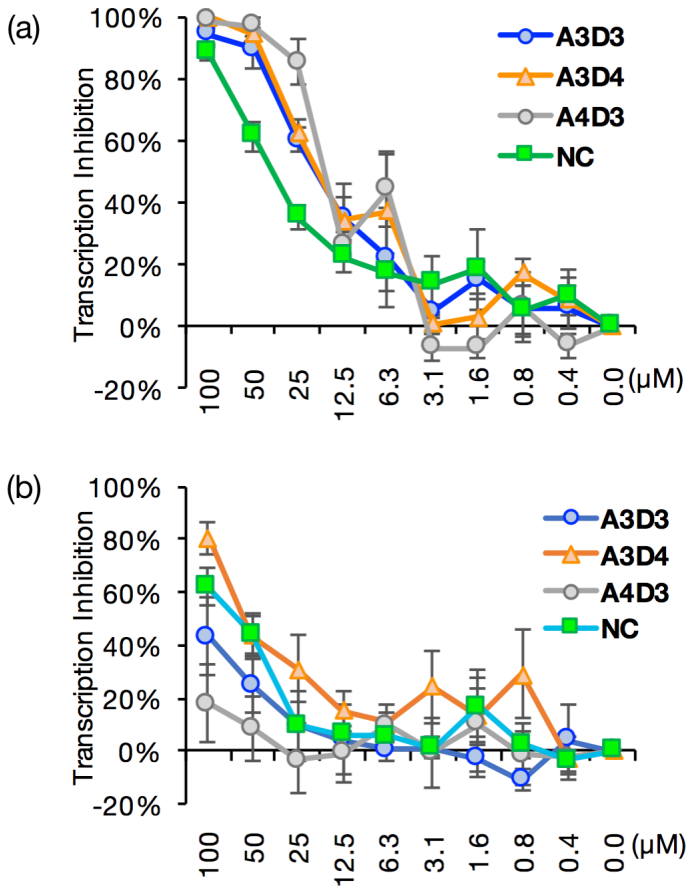


Figure 30. Plot of the percentage of *in vitro* transcription of CTG_{74} versus ligand concentrations. Ligands at different concentrations were incubated with 15 ng of linearized CTG_{74} in a T7 RNA polymerase mixture. After 1.5 h, the reaction mixture was loaded on a 10% denaturing gel. Noncyclic control **56** is **NC**.

A3D4, and **A4D3** moderately inhibited the production of r(CUG)^{exp} (~60–80% at 25 μM) in a dose-dependent manner, whereas the noncyclic control ligand **56** showed ~45% inhibition (Figure 29a, **NC**). In addition, to study the selectivity and potentially the target of the inhibition, similar experiments were performed with a control plasmid that lacked repeats. Ligand **A4D3** showed <10% transcription inhibition at 0–50 μM (<20% at 100 μM), whereas **A3D3** showed ~45% inhibition at 100 μM and **A3D4** showed ~45% inhibition at 50 μM. In contrast, the control ligand **56** showed ~45% inhibition at 50 μM (Figure 29b). Interestingly, the most effective inhibitor, **A4D3**, was the most selective inhibitor yet also a relatively toxic macrocycle. It is important to point out that the macrocycles were still able to inhibit transcription using dsDNA that did not contain mismatched base pairs, which suggests, given their minimal binding affinity toward dsDNA, that the macrocycles likely inhibit transcription through ways other than binding the d(CTG)^{exp} hairpin that is proposed to exist during transcription.[33,34,42] Taking all the results into account, macrocycle **A3D3** appears to be the overall better macrocycle due to its lower cytotoxicity; however, there does not seem to be a clear trend between structure and bioactivity. It is possible that the macrocycles could perform differently *in vivo*, and the results thus far do not reflect the complexity environment of the cell.

3.3.6. Conclusions and Outlook

Through rational design, we have developed a new class of water-soluble cyclobisintercalators that selectively target T–T mismatches and have negligible nonspecific binding toward dsDNA. They are able to penetrate through the cell membrane, are less cytotoxic, and about twice as efficient at inhibiting the transcription of d(CTG)^{exp} to r(CUG)^{exp} than the noncyclic analog, **56**. More significant was the unexpected discovery that the

macrocyclic structure design of our ligands did not necessarily correlate with a reduction in their cytotoxicity in HeLa cells. Whereas ligands **A3D3** and **A3D4** showed almost 10-fold lower toxicities compared to noncyclic control **56**, **A4D4** and **A4D3** showed a very similar toxicity profile (Figure 28). The contrasting cytotoxic profiles of A3D4 and A4D3 (IC_{50} >100 and ~20 μM, respectively) suggest that molecule size does not have a significant effect on toxicity, but rather the molecular conformations play a significant role. In addition, varying the lengths of the linkers in the macrocyclic ligands did not have a significant effect on their binding affinities toward CTG sites.

A few key questions still remain: 1) What is the nature of the cytotoxicity of the acridine-based ligands (cyclic and noncyclic)? 2) What is the mode of binding of the macrocyclic ligands to DNA? To further investigate the cytotoxicity of the acridine-based, one could take advantage of activity-based protein profiling methods, specifically tandem-orthogonal proteolysis (TOP)-activity-based protein profiling (ABPP),[127] to explore potential protein targets that could be triggering cell death. Regarding the macrocyclic ligands presented here, one could modify the synthesis to include an alkyne group that would allow the protein-macrocycle complex to be pulled down and be analyzed using mass spectrometry.

Lastly, due to the very weak binding observed of **A3D4**, **A4D4**, and **A5D5** toward C–C mismatches, it is possible that the macrocycles induce a base flip out as the C–C mismatches are one of the least stable base pair mismatches.[108] In efforts to increase the binding affinity of the macrocyclic ligands, it may be possible to utilize a different recognition unit that positioned more favorably to hydrogen-bond with a T-base. For example, one could use a triaminopyrimidine unit instead of the triaminotriazine unit to possibly position the

donors/acceptors more favorably. Alternatively, a different intercalating unit could be used to enhance the intrahelical π-stacking of the ligand in the mismatched site.

These findings will aid our design and optimization efforts of future mismatch-targeting ligands, and we are particularly interested in better understanding their cytotoxicity mode of action. Although the macrocycles in this work showed slightly better transcription inhibition of $d(CTG)_{74}$ than the noncyclic control ligand, further optimization of the design could enhance their efficacy and potentially lead to more specific transcription inhibitors and DM1 therapeutic agents.

CHAPTER 4

MATERIALS AND METHODS

4.1. General Methods

Unless otherwise stated, all solvents and reagents were of reagent quality, purchased commercially, and used without further purification. Water was purified by a Millipore Direct Q-5 purification system. DNA and RNA oligomers were purified by standard desalting and obtained from Integrated DNA Technologies. All chemical reactions were performed under nitrogen unless otherwise specified. Abbreviations used: Boc (*tert*-butyloxycarbonyl), Cbz (carboxybenzyl), DCM (dichloromethane), DIPEA (diisopropylethylamine), DME (1,2-dimethoxyethane), DMF (dimethylformamide), TEA (triethylamine).

High- and low-resolution resolution mass spectra were obtained by the Mass Spectrometry Laboratory, School of Chemical Science, University of Illinois. Mass spectra were obtained by ESI on a Waters Micromass Q-Tof. High-performance liquid chromatography (HPLC) was performed by a Dynamax SD-200 system with a UV detector set at 254 nm using an Alltech Denali C-18 column (250 x 10 mm) with a dual solvent system of (0.1:100 v/v) TFA/H_2O (Solvent A) and (0.1:100 v/v) TFA/MeOH (Solvent B). Nuclear magnetic resonance (NMR) spectra were recorded on a Varian *Unity* 400, Varian *Unity* 500 or Varian *INOVA* 500NB spectrometer at ambient temperature (22 ± 3 °C) unless otherwise mentioned. Chemical shifts (δ) are reported in parts per million (ppm). Coupling constants (*J*) were reported in Hertz. ^{1}H NMR chemical shifts were referenced to the residual solvent peak at 7.26 ppm in chloroform-*d* ($CDCl_3$) and at 2.50 ppm in DMSO-d_6. ^{13}C NMR chemical shifts were referenced to the center solvent peak at 77.16 ppm for $CDCl_3$ and at 40.45 ppm for DMSO-d_6. Unless otherwise stated, analytical thin-layer chromatography (TLC) was performed on Merck pre-coated silica

gel 60 F_{254} plates. Flash column chromatography was performed on 40-63 μm silica gel (SiO_2) mixed with Fluorescence Indicator green 254 nm in quartz columns to allow monitoring of fractions with a UV lamp. Unless otherwise stated, all purification columns were slurry-packed and wet-loaded with the crude mixture. Solvent mixtures used for chromatography are reported as volume ratio (v:v). The purity of all ligands was estimated to be >99% by ^{1}H NMR and by HPLC. Stock DNA and RNA solutions were prepared using biological grade water and THE Ambion® RNA Storage Solution (Life Technologies), respectively, and the concentrations were determined by performing absorbance measurements at 25 °C on a Shimadzu UV-2501PC spectrophotometer (Kyoto, Japan). The concentration of each single-stranded DNA/RNA was calculated using Beer's law with the extinction coefficient (ε_{260}) provided by the supplier.

Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) measurements were performed at 25 °C on a MicroCal VP-ITC (MicroCal, Inc., Northampton, MA). A standard experiment consisted of titrating 10 μL of a 500 or 700 μM ligand from a 250-μL syringe (rotating at 300 rpm) into a sample cell containing 1.42 mL of a 10 μM DNA or RNA solution. ITC experiments consisted of 28 total injections (first injection was 1–5 μL over 5 s, subsequent injections were 10 μL over 20 s), with a delay of 400 s between injections. The initial delay prior to the first injection was 300 s. To derive the heat associated with each injection, the area under each isotherm (microcalories per second versus seconds) was determined by integration by the graphing program Origin 7.0 (MicroCal, Inc. Northampton, MA). The first data point from each ITC experiment was omitted when fitting to binding models due to possible diffusive mixing of material near the tip of the syringe.[128] The fitting requirements were such that the thermodynamic parameters were derived from curves that

produced the lowest amount of deviation. In most cases, fitting to a sequential site binding model of two binding sites gave the most accurate data, and in some cases a two-site binding model fit best. The ligand stock solution was 10 mM in biological grade water. Double-stranded and hairpin DNA or RNA solutions were freshly prepared by mixing required volumes of the corresponding single stranded oligomers and annealing by heating in a water bath at >90 °C for 5 min and slowly cooling to room temperature over 60-90 min. MOPS buffer solution (1 M, pH 7 ± 0.2), NaCl solution (5 M), and biological grade water were added to make up an oligonucleotide solution with 20 mM MOPS and 300 mM NaCl. For experiments using RNA, appropriate volumes of THE Ambion® RNA Storage

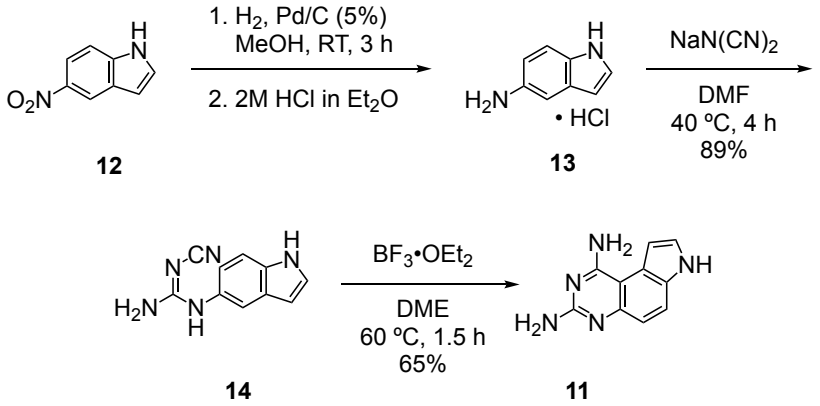
Job Plot Analysis. Fluorescence measurements were carried out at 25 °C on a HORIBA Jobin Yvon (Edison, NJ) FluoroMax-3 spectrofluorometer using a 250 μL quartz cell with 0.1 cm path length. The concentrations of the DNA and ligand were varied in each sample whereas the sum of the concentrations (c_0) was kept constant. All solution contained MOPS (20 mM; pH 7.0 ± 0.2) and NaCl (300 mM). The blanks, consisting of the same solutions with no DNA, were used to correct dilution effects. The fluorescence signal was measured with an excitation wavelength of 425 nm and emission wavelength was collected from 445 to 540 nm. The fluorescence intensity of the sample (F) and the blank (F_0) at 463 nm for each mole fraction of ligand (X_L) was used for the Job plot analysis. A Job plot was constructed by plotting the change in fluorescence ($\Delta F = F_0 - F$) versus the ligand mole fraction. The stoichiometry was determined from the intercept of the two tangents aligned at the positions $X_L = 0.0$ and 1.0.

In Vitro **Transcription.** *In vitro* transcription assays were carried out as previously reported previously with minor modifications.[75] Briefly, the ligands at each concentration were incubated with 15 ng of template DNA (linearized $(CTG)_{74}$ plasmid or control DNA template),

0.5 mM each rNTPs, and 1X T7 transcription buffer (Biolab, 80 mM Tris pH 8.3, 10 mM $MgCl_2$, 2 mM spermine, 0.1% Triton-X, 10 mM NaCl) at 37 °C for 3 h. After pre-incubation, 0.5 U T7 RNA polymerase (Biolab) was added to the reaction mixture, and transcription was allowed for 1.5 h. The reaction was quenched by adding 7 μL of 8 M urea, 1 μL of denaturing dye (95% formamide, 5 mM EDTA, 0.025% each xylene cyanol and bromophenol blue) and 2 μL of 10 μM HIV fs RNA, and then heating to 95 °C for 5 min. Of this solution, 15 μL was run on a 10% denaturing polyacrylamide gel in 1× TBE. The gel was stained with EtBr. Bands were quantified using ImageJ (NIH).

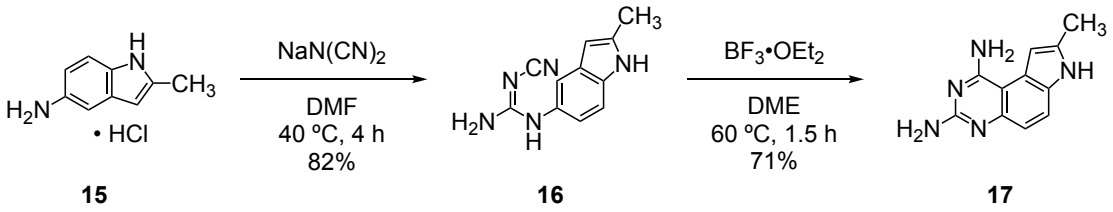
Cytotoxicity Study. Cytotoxicity of compounds was studied using HeLa cells by following the protocol reported by Vichai and Kirtikara.[126] Briefly, HeLa cells were prepared in a 96-well plate a day before the compound treatment. The ligand was added to each well to have serially diluted concentrations from 100 μM to 0.2 μM and incubated at 37°C for 72 h. Then, the cells were fixed and stained with sulforhodamine B (SRB). Bound SRB was dissolved by Tris base solution (pH 10.5) and the OD was measured at 510 nm. The percentage of cell death was plotted against concentrations of compounds.

4.2. Synthetic Procedures

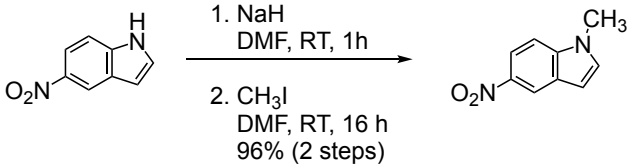


12 **13** **14** **11**

Ligand 11. $NaN(CN)_2$ (3.65 g, 40.8 mmol) was added to a stirred solution of **13** (2.75 g, 16.3 mmol, prepared by treating a methanolic solution of 5-aminoindole with 1.5 equiv. HCl in Et_2O) in DMF (27 mL) in a 50-mL round-bottomed flask. The reaction mixture was stirred at 40 ºC for 4 h. DMF was removed and the residue was treated with H_2O (50 mL) vigorously overnight. A gray solid precipitated and was collected by filtration and dried under high vacuum to give compound **14** (2.63 g, 89% yield), which was used for the next step without further purification. The characterization data were consistent with literature reported values:[88] ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.12 (s, 1 H), 8.86 (s, 1 H), 7.46 (s, 1 H), 7.35-7.33 (m, 2 H), 6.94 (dd, J = 8.8 Hz, 2.0 Hz, 1 H), 6.74 (s, 2 H), 6.40 (s, 1 H). Boron trifluoride (9.4 mL, 76 mmol) was added dropwise to a stirred suspension of **14** (2.63 g, 13.2 mmol) in DME (300 mL) and stirred at 60 ºC for 4 h. The solvent was removed, and the residue was suspended in MeOH (30 mL) and treated with NH_4OH (40 mL) for 2 h. The solvents were removed in *vacuo* and the residue was purified by column chromatography on 125 mL Al_2O_3 (activated, basic), eluting with gradient 5:1 to 1:1 (over 3 L) DCM:MeOH to afford 1.9 g (65%) of ligand **11** as a white to pale yellow solid. ^{1}H NMR (400 MHz, DMSO-*d*6) δ 11.55 (s, 1 H), 7.64 (d, J = 8.8 Hz, 1 H), 7.43 (t, J = 2.8 Hz, 1 H), 7.03-7.00 (m, 2 H), 6.65 (brs, 2 H), 5.65 (s, 2 H); ^{13}C NMR (100 MHz, DMSO-*d*6) δ 162.0, 159.1, 150.1, 130.4, 124.8, 120.0, 119.1, 119.0, 102.5, 102.0.



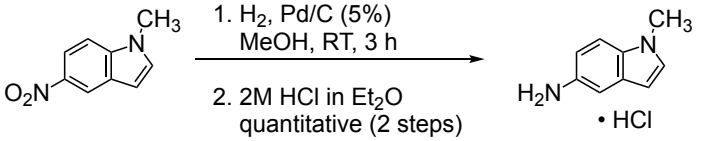
Ligand 17. $NaN(CN)_2$ (1.52 g, 17.1 mmol) was added to a stirred solution of **15** (1.25 g, 6.84 mmol, prepared by treating a methanolic solution of commercially-available 2-methyl-1*H*-indol-5-amine with 1.5 equiv. HCl in Et_2O) in DMF (12 mL) in a 50-mL round-bottomed flask. The reaction mixture was stirred at 40 ºC for 4 h. DMF was removed and the residue was treated with H_2O (20 mL), sonicated briefly, and left stirring vigorously overnight. A gray solid precipitated and was collected by filtration and dried under high vacuum to give compound **16** (1.19 g, 82% yield), which was used for the next step without further purification. ^{1}H NMR (400 MHz, DMSO-d_6) δ 10.92 (s, 1H), 8.82 (s, 1H), 7.30 (d, J = 2.0 Hz, 1H), 7.20 (d, J = 8.5 Hz, 1H), 6.84 (dd, J = 8.5, 2.1 Hz, 1H), 6.72 (s, 2H), 6.09 (s, 1H), 2.36 (s, 3H). Boron trifluoride etherate (3.5 mL, 28.5 mmol) was added dropwise to a stirred suspension of **16** (1.27 g, 5.94 mmol) in DME (115 mL) and stirred at 60 ºC for 4 h. The solvent was removed, dried overnight, and the residue was suspended in MeOH (12 mL) and treated with NH_4OH (8 mL) for 2 h. The solvents were removed *in vacuo* and the residue was purified by column chromatography on 120 mL Al_2O_3 (activated, basic), eluting with gradient 7:1 to 2:1 (over 2 L) DCM:MeOH to afford 0.89 g (71%) of ligand **17** as a pale yellow solid. ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.29 (s, 1H), 7.52 (d, J = 8.8 Hz, 1H), 6.92 (d, J = 8.8 Hz, 1H), 6.73 (s, 1H), 6.52 (s, 2H), 5.56 (s, 2H), 2.43 (s, 3H).



18 **19**

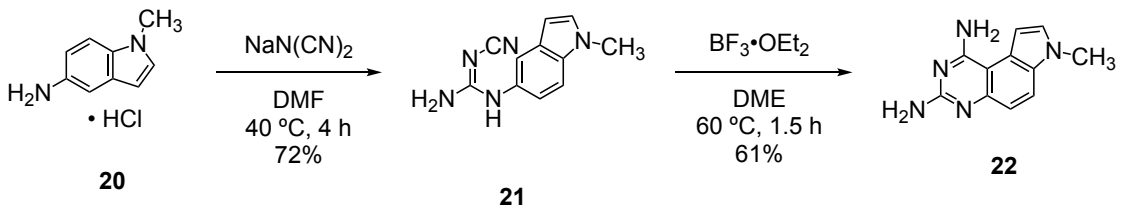
1-Methyl-5-nitro-1*H*-indole (19). To a stirred solution of **18** (1.0 g, 6.16 mmol) in 20 mL of anhydrous DMF in a 100-mL round-bottomed flask was added sodium hydride (60% in mineral oil, 0.27 g, 6.78 mmol). The mixture was stirred at r.t. for 1 h and then iodomethane

(0.42 mL, 6.78 mmol) was added and stirred at r.t. overnight, monitoring by TLC (25:1 DCM:MeOH, R_f = 0.5). The reaction was quenched with saturated aqueous ammonium chloride (35 mL) and extracted with EtOAc (2 x 40 mL). The organic layer was washed with 5% aqueous lithium chloride (50 mL), H_2O (50 mL), brine (100 mL), and dried with $MgSO_4$ and concentrated *in vacuo* to afford 1.04 g of **19** (96%) as a dark yellow powder. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.59 (d, *J* = 2.0 Hz, 1 H), 8.13 (dd, *J* = 2.0, 9.0 Hz, 1 H), 7.34 (d, *J* = 9.0 Hz, 1 H), 7.20 (d, *J* = 3.0 Hz, 1 H), 6.67 (dd, *J* = 0.5, 1.0, 3.0, 3.5 Hz, 1 H), 3.86 (s, 3 H).



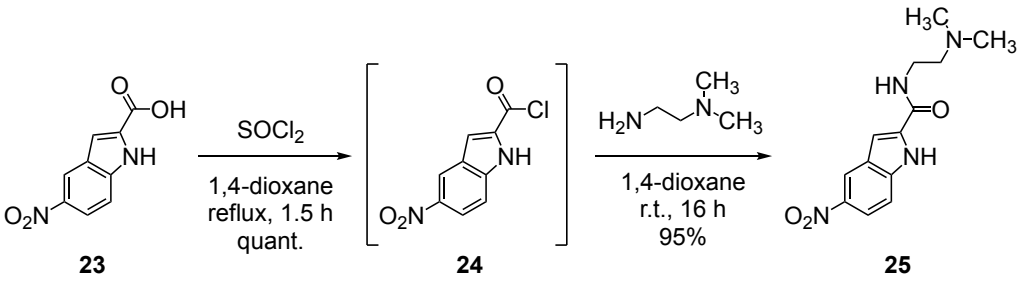
19 **20**

1-Methyl-1*H*-indol-5-amine (20). To a 100-mL round-bottomed flask containing 0.2 g of Pd/C (5%) carefully wetted with ca. 3 mL of methanol and a magnetic stir bar under an inert atmosphere was added a solution of 0.5 g (2.8 mmol) of **19** in 40 mL of methanol. Hydrogen gas was bubbled through the solution while stirring at r.t. for 1 h. The solution was vacuum-filtered through a small Celite bed and the solvents were removed *in vacuo* to afford 0.41 g (quant.) of **20** as a brown powder. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.12 (dd, *J* = 12.1, 7.5 Hz, 2 H), 6.84 (t, *J* = 1.6 Hz, 1 H), 6.77 (dd, *J* = 7.6, 1.5 Hz, 1 H), 6.54 (dd, *J* = 7.5, 1.5 Hz, 1 H), 3.73 (d, *J* = 10.1 Hz, 5 H).



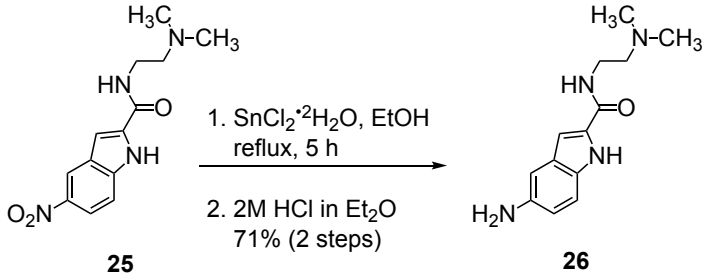
20 **21** **22**

Ligand 22. The title compound was prepared as described for ligand **11** from 0.31 g (2.1mmol) of **20**, affording 0.32 g (72%) of **21** and 0.19 g (61%) of ligand **22** as a yellowish solid. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.11 – 8.03 (m, 2 H), 7.37 (d, J = 7.6 Hz, 1 H), 7.02 (d, J = 7.5 Hz, 1 H), 5.38 (s, 2 H), 4.46 (s, 2 H).

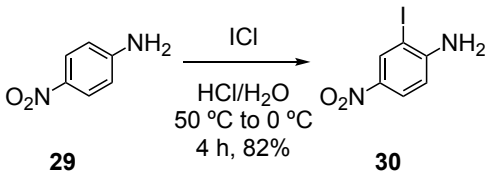


***N*-(2-(Dimethylamino)ethyl)-5-nitro-1*H*-indole-2-carboxamide (25).** To a stirred solution of 0.5 g (2.43 mmol) of **23** in 6 mL of dioxane in a 25-mL round-bottomed flask was slowly added 6 mL of freshly distilled $SOCl_2$. The mixture was refluxed at 105 ºC for 1.5 h and then the excess $SOCl_2$ was distilled off; the last traces were removed azeotropically via co-evaporation with DCM (4 x 15 mL). The crude was left under high vacuum for 1 h to afford the crude intermediate **24** as a dark yellow powder. Then it was dissolved in 7 mL of dioxane at r.t. and 1 mL (9.22 mmol) of *N*,*N*-dimethylethylenediamine was added. The mixture was stirred at RT overnight, monitoring by TLC on aluminum oxide (activated, basic) plates (25:1 DCM:MeOH, R_f = 0.5). The solvents were removed, and the remaining crude solid was dissolved in EtOAc/H_2O (200 mL) mixture and extracted with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (100 mL), dried with $MgSO_4$ and concentrated *in vacuo* and the crude was purified by column chromatography (~120 mL Al_2O_3, activated, basic) using 300:1-300:3 gradient eluent of DCM:MeOH. The product- containing fractions were combined and concentrated *in vacuo* to afford 0.64 g (95%) of **25** as a light-yellow powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.29 (s, 1 H), 8.70 (d, J = 2.2 Hz, 1 H),

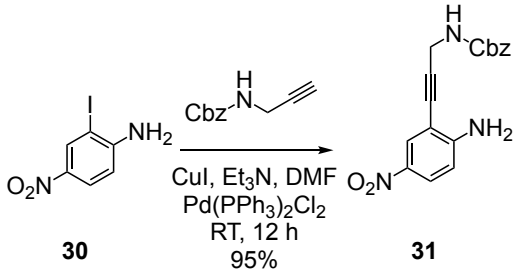
8.66 (t, J = 5.7 Hz, 1 H), 8.06 (dd, J = 9.1, 2.3 Hz, 1 H), 7.57 (d, J = 9.1 Hz, 1 H), 7.38 (d, J = 0.9 Hz, 1 H), 3.39 (q, J = 6.5 Hz, 3 H), 2.42 (t, J = 6.7 Hz, 2 H), 2.19 (s, 6 H).



5-Amino-*N*-(2-(dimethylamino)ethyl)-1*H*-indole-2-carboxamide (26). To a stirred solution of 0.46 g (1.66 mmol) of **25** in 25 mL of EtOH was added 3.75 g (16.6 mmol) of $SnCl_2$•$2H_2O$ and refluxed (100 ºC) for 24 h, monitoring by TLC on aluminum oxide (activated, basic) plates (25:1 DCM:MeOH, R_f = 0.3). The solvents were removed and the residue was dissolved in 150 mL of water and basified to pH 8 with aqueous $NaHCO_3$ or 2M KOH. This was extracted with EtOAc (6 x 150 mL). The combined organic layers were washed with brine (100 mL), dried with $MgSO_4$ and concentrated *in vacuo* and the crude was purified by column chromatography (~120 mL Al_2O_3, activated, basic) using 300:2 DCM:MeOH (MeOH contained 5% NH_4OH v/v).The product-containing fractions were combined and concentrated *in vacuo* to afford 0.30 g (74%) of **26** as a light brown powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.67 (s, 1 H), 7.33 (d, J = 7.6 Hz, 1 H), 6.99 (d, J = 1.6 Hz, 1 H), 6.82 (t, J = 1.4 Hz, 1 H), 6.48 (dd, J = 7.5, 1.7 Hz, 1 H), 4.26 (s, 2 H), 3.35 (d, J = 7.2 Hz, 2 H), 2.43 (t, J = 7.3 Hz, 2 H), 2.20 (s, 6 H).

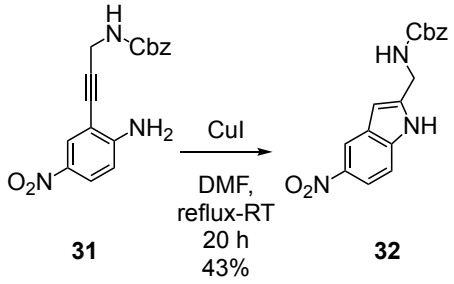


2-Iodo-4-nitroaniline (30). (Prepared from a reported procedure24 with slight modifications.) To a stirred solution of 1.38 g (10 mmol) of **29** in 4 mL of concentrated HCl and 30 mL of H_2O in a 200-mL round-bottomed flask was added a solution of 3.41 g (21 mmol) of ICl in 6 mL of concentrated HCl drop-wise over 1 h using a syringe pump. The mixture was stirred at 50 ºC for 2.5 h, then at 0 ºC for 15 min. Two grams of Na_2SO_4 were added, and stirred at RT for 15 min. Diethyl ether (30 mL) was added and stirred for an additional 20 min. The aqueous layer was extracted with diethyl ether (2 x 70 mL), and the organic layers were washed once with $NaHCO_3$ (50 mL), dried with $MgSO_4$. The organic layers were concentrated and purified by column chromatography (~100 mL SiO_2) using 15:1 hexanes:ethyl acetate, and monitored by TLC (5:1 hexanes:EtOAc, R_f = 0.30). The product-containing fractions were combined and concentrated *in vacuo* to afford 2.16 g (82%) of **30** as a brown crystalline solid.[1]H HMR (400 MHz, $CDCl_3$) 8.57 (d, J = 3Hz, 1H), 8.10 (dd, J = 3, 9Hz, 1H), 6.47 (d, J = 9 Hz, 1H), 4.80 (br d, 1H).



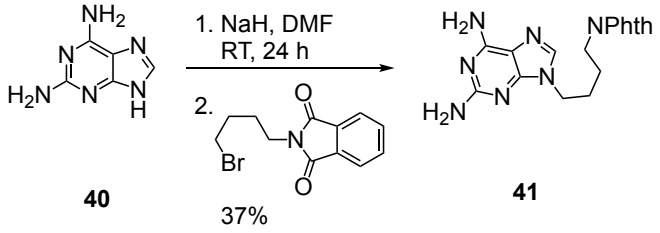
Benzyl (3-(2-amino-5-nitrophenyl)prop-2-yn-1-yl)carbamate (31). To a stirred solution of 0.505 g (1.91 mmol) of **30**, 5.46 g (0.029 mmol) of CuI and 20.1 mg (0.029 mmol) of $Pd(Ph_3P)_2Cl_2$ in 22 mL of TEA (22 mL) was added 1.30 g (6.86 mmol) of Cbz-protected propargyl amine drop-wise. The mixture was stirred at r.t. for 12 h, monitoring by TLC (5:1 hexanes:EtOAc, R_f = 0.12). The solvents were removed *in vacuo* and the crude was purified

by column chromatography (~100 mL SiO_2) using an eluent gradient of 20:1 to 10:1 hexanes:ethyl acetate over 1.5 L. The product-containing fractions were combined and concentrated *in vacuo* to afford 0.59 g (95%) of **31** as a dark orange solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.18 (d, J = 2.7 Hz, 1 H), 8.03 (dd, J = 9.0, 2.6 Hz, 1 H), 7.39 (d, J = 4.6 Hz, 5 H), 6.65 (d, J = 9.1 Hz, 1H), 5.17 (s, 2 H), 5.03 (s, 2 H), 4.26 (d, J = 5.8 Hz, 2 H), 2.19 (s, 1H).



31 **32**

Benzyl ((5-nitro-1*H*-indol-2-yl)methyl)carbamate (32). To a stirred solution of 0.606 g (1.86 mmol) of **31** in 8 mL of DMF was added 0.12 g (0.612 mmol) of CuI. The mixture was stirred at reflux for 4 h, then at r.t. for 16 h, monitoring the reaction by TLC (3:2 ethyl acetate:hexanes, R_f = 0.5). The solvent was removed *in vacuo* to leave a dark residue, which was purified by column chromatography (~100 mL SiO_2) using an eluent 25:1 hexanes:ethyl acetate to afford 0.26 g (43%) of **32** as a dark orange oil. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.78 (t, J = 1.6 Hz, 1 H), 7.94 – 7.85 (m, 2 H), 7.71 (d, J = 7.5 Hz, 1 H), 7.38 – 7.27 (m, 2 H), 6.55 (d, J = 1.5 Hz, 1 H), 5.45 (s, 1 H), 5.03 (s, 2 H), 4.75 (s, 1 H), 4.46 (s, 1 H).

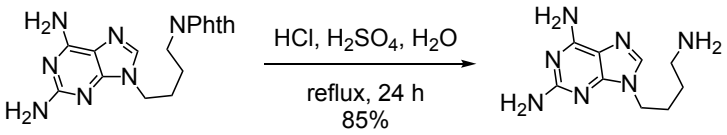


40 **41**

37%

2-(4-(2,6-Diamino-9*H*-purin-9-yl)butyl)isoindoline-1,3-dione (41). To a stirred solution of 0.514 g (3.42 mmol) of 2,6-diaminopurine in 15 mL of anhydrous DMF in a 50-mL (24/40)

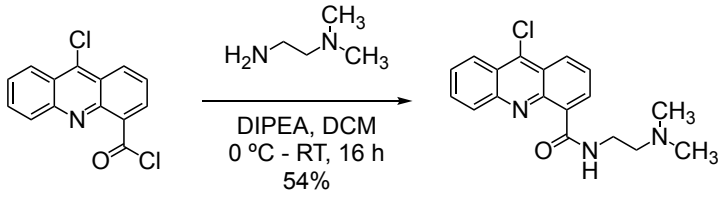
round-bottomed flask was added 0.155 g (3.86 mmol) of a 60% suspension in oil of NaH and stirred at RT for 2 h. Then, 0.966 g (3.42 mmol) of *N*-(4-bromobutyl)phthalimide was added, and the mixture was stirred for 24 h. The reaction was monitored by TLC (silica gel plates, 10:1 (v/v) DCM:MeOH, $R_f = 0.4$). The solvents were removed *in vacuo*, the remaining crude was sonicated in 1:1 (v/v) MeOH:H_2O, and the solid was filtered off and washed with MeOH thoroughly, affording 0.443 g (1.26 mmol, 37%) of **41** as an off-white powder. ^{1}H NMR (500 MHz, DMSO-d_6) δ 7.89 – 7.79 (m, 4H), 7.64 (d, J = 1.6 Hz, 1H), 6.62 (s, 2H), 5.75 (s, 2H), 3.96 (t, J = 7.0 Hz, 2H), 3.59 (t, J = 7.2 Hz, 2H), 1.74 (p, J = 6.9 Hz, 2H), 1.54 (p, J = 7.2 Hz, 2H).



41 **42**

9-(4-Aminobutyl)-9*H*-purine-2,6-diamine (42). A solution of 0.443 g (1.26 mmol) of **41** in 5 mL of a 1:1:1 (v/v) mixture of H_2O:AcOH:HCl (conc.) in a 50-mL (24/40) round-bottomed flask equipped with a condenser was stirred at 100 ºC for 24 h. The reaction was monitored by TLC (silica gel plates, 10:1 (v/v) DCM:MeOH). Then, some of the solvents were removed *in vacuo*, producing a white precipitate (phthalic acid). This precipitate was filtered off and the remaining solvents were removed *in vacuo*. The crude was dissolved in water, and the solvents were removed *in vacuo* (this was repeated two times). The crude was dissolved in minimal water (*ca*. 7 mL), basified with 3M KOH, and then saturated with NaCl powder. The resulting suspension was transferred to a separatory funnel, and extracted multiple times with 10:1 (v/v) CH_3Cl:*n*-BuOH. The organic layers were combined, dried over $MgSO_4$, and concentrated *in*

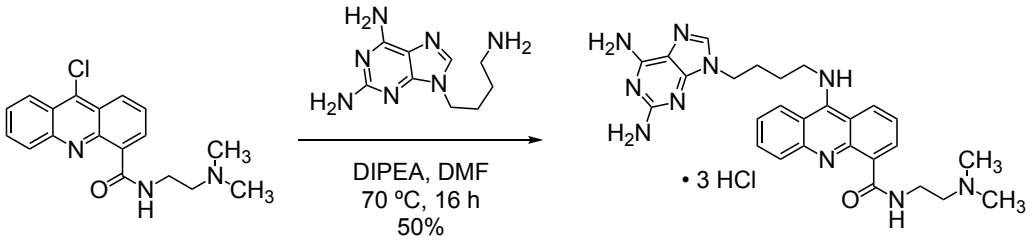
vacuo, affording 0.237 g (1.07 mmol, 85%) **42** as a clear oil. ^{1}H NMR (500 MHz, DMSO-d_6) δ 7.62 (d, J = 1.6 Hz, 1H), 6.63 (s, 2H), 5.72 (s, 2H), 3.94 (t, J = 7.0 Hz, 2H), 2.69 (t, J = 7.2 Hz, 2H), 1.74 (p, J = 6.9 Hz, 2H), 1.47 (p, J = 7.2 Hz, 2H).



43 **44**

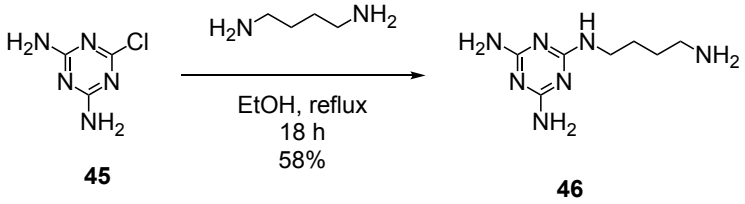
9-Chloro-*N*-(2-(dimethylamino)ethyl)acridine-4-carboxamide (44). To the 65-mL (14/20) round-bottomed flask (sealed with a septum and equipped with a magnetic stir bar) containing crude **43** under argon (4.06 mmol), which was carefully collected avoiding exposure to air and moisture as much as possible, was added 4 mL of anhydrous DCM *via* a syringe and needle, followed by 1 mL of DIPEA and then quickly placed in an ice bath. Additional DIPEA (*ca*. 0.8 mL) was added to ensure the reaction remained at a basic pH. *N*,*N*-Dimethylethylenediamine (0.256 g, 2.91 mmol) was dissolved in 4 mL of anhydrous DCM and was cannulated into the round-bottomed flask at once, and its glass vial was rinsed with 4 additional mL of anhydrous DCM and also transferred to the round-bottomed flask. The resulting mixture was periodically sonicated to break apart some of the solid precipitates. The mixture was stirred at 4 ºC for 1 h, then at room temperature overnight, resulting in a dark orange/yellow turbid solution. The reaction was monitored by TLC using silica gel plates and a mixture of 10:1 DCM:MeOH. The mixture was diluted with 5 mL of methanol, vacuum filtered through a Celite pad and rinsed with 15 mL of methanol, and the filtrate was concentrated *in vacuo* leaving the crude as a dark orange solid. The crude was dissolved in *ca*. 5 mL of 100:1 (v/v) DCM:MeOH and loaded onto a silica gel column packed with the same

solvent conditions (column was 4.5 cm in diameter, 26 cm in height). The column was eluted with a solvent gradient of 500:5 to 500:40 DCM:MeOH over 3 L, collecting 200-mL fractions (product started eluting after *ca*. 1.5 L) and monitoring by TLC (R_f = 0.48). The fractions containing the product were concentrated and repurified using the same conditions above, affording 0.515 g (1.57 mmol, 54%) of **44** as a dark yellow solid. ^{1}H NMR (500 MHz, Chloroform-*d*) δ 11.44 (s, 1H), 9.00 (dd, *J* = 7.1, 1.5 Hz, 1H), 8.54 (dd, *J* = 8.7, 1.5 Hz, 1H), 8.39 (dd, *J* = 8.7, 1.3 Hz, 1H), 8.11 (dd, *J* = 8.7, 1.0 Hz, 1H), 7.85 (ddd, *J* = 8.3, 6.6, 1.3 Hz, 1H), 7.77 – 7.62 (m, 2H), 3.61 (t, *J* = 5.1 Hz, 2H), 2.47 (t, *J* = 5.1 Hz, 2H), 2.29 (s, 6H).

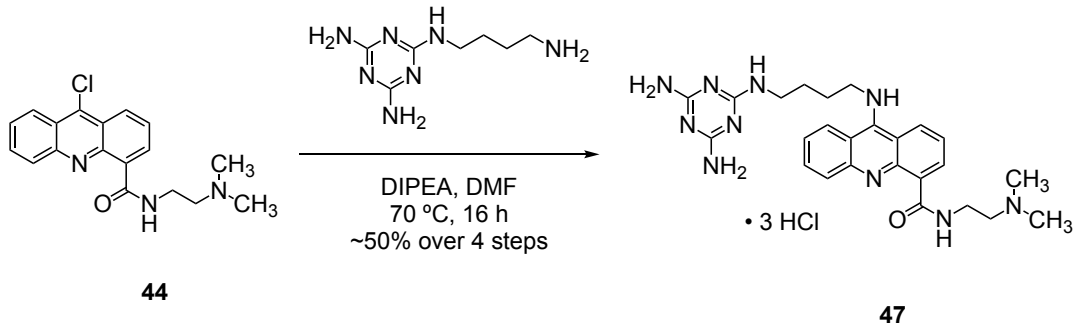


Ligand 39. To a stirred solution of 0.233 g (0.71 mmol) of **44** in 5 mL of anhydrous DMF in a 25-mL (14/20) round-bottomed flask was added 0.2 mL of DIPEA and 0.137 g (0.62 mmol) of **42** and stirred at 70 ºC for 16 h. The reaction was monitored by TLC (Al_2O_3, 10:1 (v/v) DCM:MeOH, R_f = 0.5). The solvents were removed *in vacuo* and the crude was further dried under high vacuum for a few hours, leaving a thick orange/brown oily residue. The residue was suspended in 10 mL of water and 6M aq. HCl until the pH was acidic, turning to a dark orange solution with some precipitate. The precipitate was filtered off through a syringe filter (0.2 micron) and loaded onto a C_{18}-reversed-phase silica gel column that was primed with 550 mL of 500:50:0.5 (v/v) H_2O:MeOH:HCl (conc.). The column was eluted with a solvent gradient of 500:50:0.5 to 500:200:0.7 H_2O:MeOH:HCl (conc.). To monitor by TLC, 6 drops

of the corresponding fraction were mixed with 1-2 drops of MeOH:NH_4OH (9:1, v/v) and the solution was spotted onto an Al_2O_3 TLC plate (basic) and developed with a mixture of 10:1 DCM:MeOH (R_f = 0.5). The fractions containing the product were combined and concentrated *in vacuo* to afford 0.193 g (0.31 mmol, 50%) of **39** as a yellow solid. ^{1}H NMR (500 MHz, DMSO-d_6) δ 13.52 (s, 1H), 12.36 (s, 0H), 10.50 (s, 1H), 10.25 (s, 0H), 9.66 (s, 1H), 8.66 (d, J = 7.4 Hz, 0H), 8.50 (s, 1H), 8.11 – 7.96 (m, 2H), 7.63 (s, 1H), 7.57 (s, 1H), 7.45 (s, 1H), 4.11 (s, 1H), 3.78 (q, J = 5.7 Hz, 1H), 3.39 (s, 0H), 2.88 (d, J = 4.4 Hz, 3H), 1.91 (s, 2H).

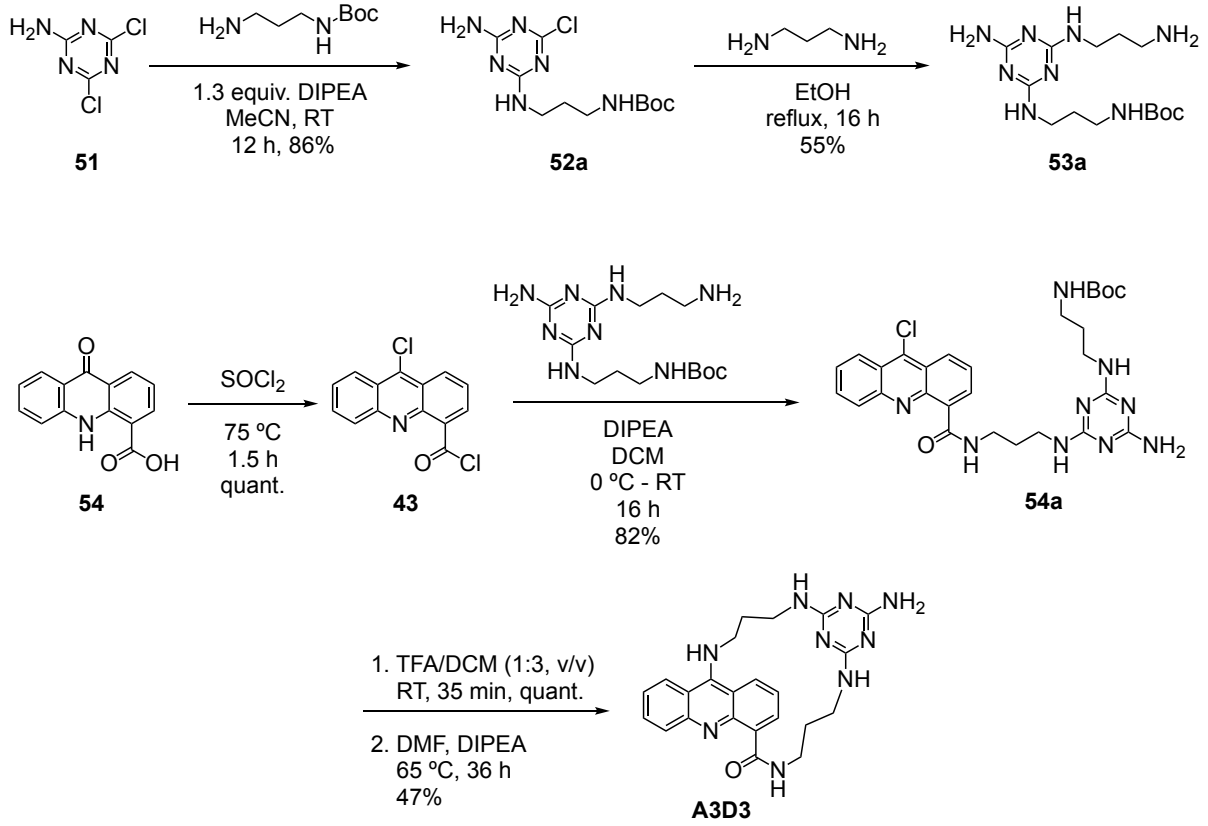


Ligand 46.[62] To a stirred solution of 7.05 g (80 mmol) of 1,4-diaminobutane at 110 ºC in a two-necked 300-mL (24/40) round-bottomed flask was added 1.05 g (7.21 mmol) of **45** (dissolved in 100 mL of EtOH) dropwise over 4 h. The mixture was refluxed for 18 h, and then the solvents were removed *in vacuo* and the residue was further dried under high vacuum overnight. The white crude solid was adsorbed onto silica and dry-loaded onto a silica column, packed with 100:12:1 (v/v) DCM:MeOH:NH_4OH. The column was eluted on a gradient up to 100:20:2 (v/v) DCM:MeOH:NH_4OH under 3 psi and checked by TLC (silica gel plates, 9:1 (v/v) MeOH:NH_4OH, R_f = 0.4).The fractions containing the product were collected and concentrated, affording 0.825 g (4.18 mmol, 58%) of **46** as a white solid. ^{1}H NMR (500 MHz; DMSO-d_6) 7.89 (s, 2H), 6.48 (t, J = 5.8, 1H), 6.00 (d, J = 80.1, 4H), 3.18 (q, J = 6.1, 2H), 2.76 (t, J = 7.4, 2H), 1.51 (ddtd, J = 25.2, 15.6, 8.4, 7.1, 4H)



Ligand 47. The title compound was prepared in similar conditions as described for compound **39** from 0.302 g (1.12 mmol) of **44** and 0.201 g (1.02 mmol) of **46**, affording 0.336 g (0.562 mmol, 55%) of **44** as a yellow solid.

MACROCYCLE A3D3



Compound 52a. To a 200-mL (24/40) round-bottomed flask equipped with a magnetic stir bar was added **51**[124] (2.05 g, 11.8 mmol), 2.1 mL (12 mmol) of DIPEA, and 55 mL of MeCN. While stirring at room temperature, *tert*-butyl(3-aminopropyl)carbamate (2.05 g, 11.8 mmol) was added and the solution was stirred for 12 h. The solution was initially colorless and white precipitate began to form after a few hours. The precipitate was filtered off and rinsed with water (30 mL) and dried under nitrogen to afford 3.12 g (10.1 mmol, 86%) of **52a** as a white solid.

Compound 53a. To a 250-mL (24/40) 2-necked round-bottomed flask equipped with a magnetic stir bar and a condenser was added 8.5 mL (102.4 mmol) of 1,3-diaminopropane and heated to 110 ºC. Compound **52a** (3.10 g, 10.2 mmol) was dissolved in 100 mL of warm ethanol and was added dropwise *via* a syringe pump at ~0.5 mL/min, eventually leading to a

refluxing mixture. After refluxing for 16 h, the solvents were removed *in vacuo* and the crude was adsorbed onto silica gel and dry-loaded on a silica gel column slurry-packed with 100:10:1 DCM:MeOH:NH_4OH (column was 4.5 cm in diameter, 28 cm in height), eluted on a solvent gradient from 100:10:1 to 100:20:2 of DCM:MeOH:NH_4OH over 2 L and monitoring by TLC (silica gel plates, using 9:1 MeOH:NH_4OH, R_f = 0.45). The fractions containing the product were collected and concentrated, affording 1.91 g (5.61 mmol, 55%) of **53a** as an off-white solid.

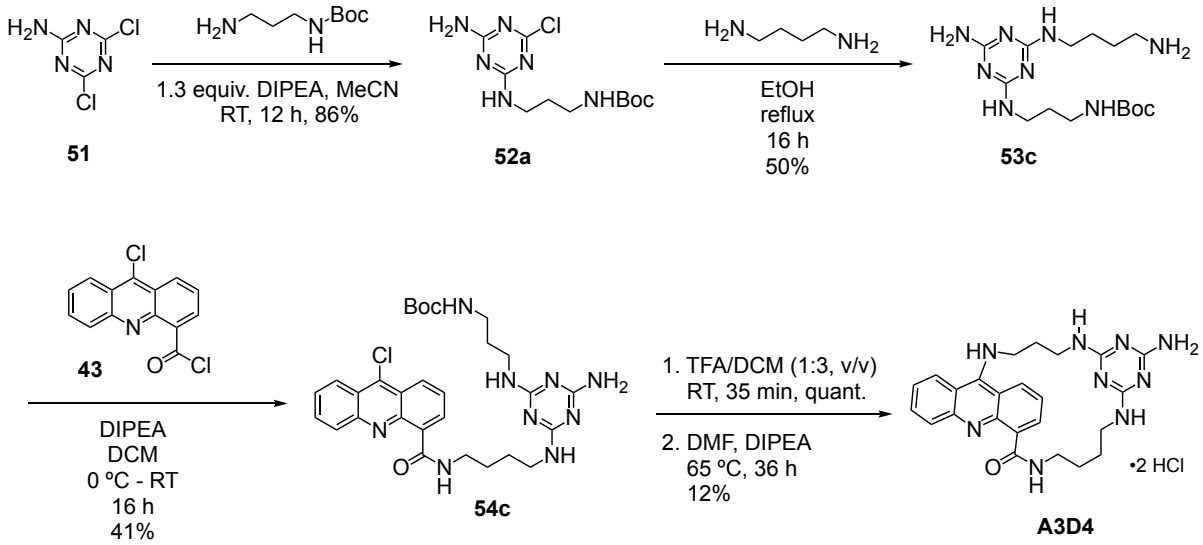
Compound 54a. To a 65 mL (14/20) round-bottomed flask (sealed with a septum and equipped with a magnetic stir bar) containing crude **43** under argon (4.68 mmol, prepared from 1.12 g of **54**[64]), which was carefully collected avoiding exposure to air and moisture as much as possible, was added 4 mL of anhydrous DCM *via* a syringe and needle, followed by 1 mL of DIPEA and then quickly placed in an ice bath. Compound **53a** (1.00 g, 2.94 mmol) was dissolved in 4 mL of anhydrous DCM and was cannulated into the round-bottomed flask at once, and its glass vial was rinsed with 4 additional mL of anhydrous DCM and also transferred to the round-bottomed flask. An additional 0.8 mL of DIPEA was added to ensure the reaction remained at a basic pH. The resulting mixture was periodically sonicated to break apart some of the solid precipitates. The mixture was stirred at 4 ºC for 1 h, then at room temperature overnight, resulting in a dark orange/yellow turbid solution. The reaction was monitored by TLC using silica gel plates and a mixture of 10:1 DCM:MeOH. The mixture was diluted with 5 mL of methanol, vacuum filtered through a Celite pad and rinsed with 15 mL of methanol, and the filtrate was concentrated *in vacuo* leaving the crude as a dark orange solid. The crude was dissolved in ca. 5 mL of 100:3.5 DCM:MeOH and loaded onto a silica gel column packed with the same solvent conditions (column was 4.5 cm in diameter, 26 cm in height). The

column was eluted with a solvent gradient of 500:17.5 to 500:40 DCM:MeOH over 3 L, collecting 200-mL fractions (product started eluting after ca. 1.5 L) and monitoring by TLC (R_f = 0.38). The fractions containing the product were concentrated and repurified using the same conditions above, affording 1.40 g (2.41 mmol, 82%) of **54a** as a dark yellow solid.

Macrocycle A3D3. To a 100-mL (24/40) oven-dried round-bottomed flask equipped with a magnetic stir bar and sealed with a rubber septum was added **54a** (0.387 g, 0.67 mmol) and kept under high vacuum for 1 h, then filled with nitrogen. Then, 25 mL of anhydrous DCM was added *via* a cannula and, while stirring vigorously at room temperature, 10 mL of TFA was added at once. The bright yellow solution was stirred for 35 min (completion of the deprotection was verified by ESI-MS) and then the solvents were removed *in vacuo*, leaving a dark yellow/orange oily residue that was further dried in high vacuum for a few hours. The crude was dissolved in ca. 6 mL of anhydrous DMF and 0.7 mL of anhydrous DIPEA was quickly added and stirred for 5–10 min (0.5 mL was needed to reach pH 9, 0.2 mL added for the reaction). The dark yellow solution was transferred to a 1-L (24/40) oven-dried round-bottomed flask equipped with a large magnetic stir bar *via* a cannula, and then ca. 550 mL of anhydrous DMF was further added *via* a cannula (final [**54a**] ≈ 1 mM). The light-yellow solution (bright blue solution under 365 nm light) was stirred at 65 ºC for 36 h, eventually turning to a deep yellow color (very bright green solution under 365 nm light). The cyclization was very sensitive to solvent dryness, easily failing if the solvent was not dry, and to the solution pH before dilution (very important that it was basic). The solvents were removed *in vacuo* and the crude was further dried under high vacuum for a few hours, leaving a thick orange/brown oily residue. The residue was suspended in ca. 20 mL of water and 6M aq. HCl was added until pH < 2, resulting in a dark orange suspension. The precipitate was filtered off

through a KimWipe-filled glass Pasteur pipette and the resulting dark orange solution was loaded onto a C_{18}-reversed-phase silica gel column that was primed with 550 mL of 500:50:0.5 (v/v) H_2O:MeOH:HCl (conc.). The column was eluted with a solvent gradient of 500:50:0.5 to 500:200:0.7 H_2O:MeOH:HCl (conc.) for 3 L. To monitor by TLC, a few drops of the corresponding fraction were mixed with 1–2 drops of MeOH:NH_4OH (9:1, v/v) and the solution was spotted onto an Al_2O_3 TLC plate (basic) using a freshly hand-custom-made capillary spotter from a Pasteur pipette and developed with a mixture of 10:1 DCM:MeOH (R_f = 0.5). The fractions containing the product were combined and concentrated *in vacuo* to afford 0.16 g (0.31 mmol, 47%) of **A3D3** as a yellow solid.

MACROCYCLE A3D4

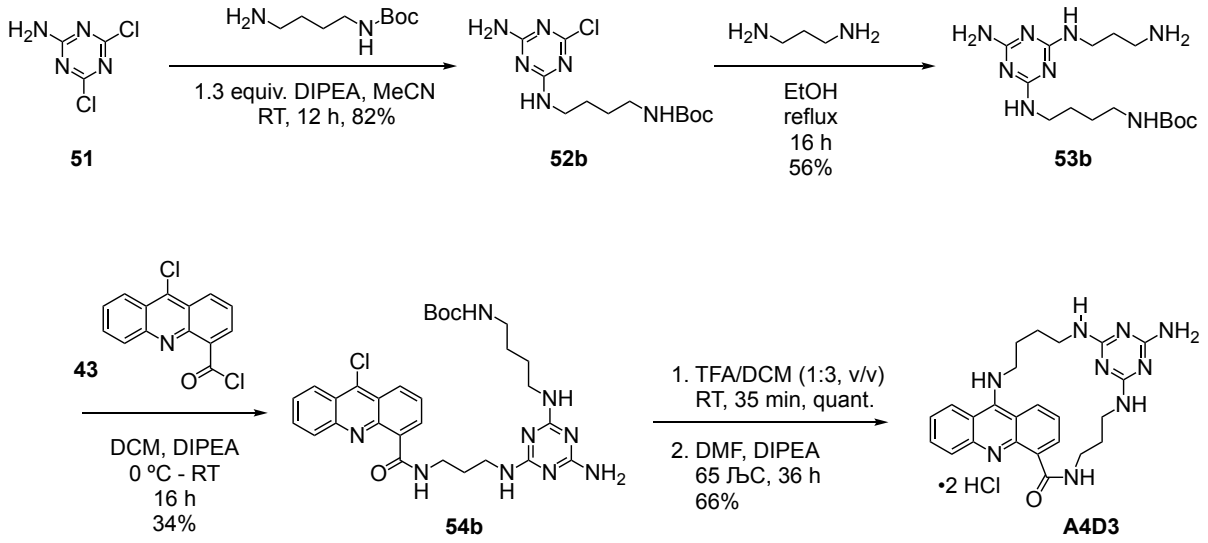


Compound 53c. The title compound was prepared in similar conditions as described for compound **53a** from 1.05 g (3.32 mmol) of **52a** and 4 mL (40 mmol) of 1,4-diaminobutane, affording 0.611 g (1.7 mmol, 50%) of **53c** as an off-white solid.

Compound 54c. The title compound was prepared in similar conditions as described for compound **54a** from 1.055 g (4.41 mmol) of **43** and 0.978 g (2.76 mmol) of **53c**, affording 0.672 g (1.13 mmol, 41%) of **54c** as a dark yellow solid.

Macrocycle A3D4. The title compound was prepared in similar as described for macrocycle **A3D3** from 0.37 g (0.62 mmol) of **54c**, affording 0.041 g (0.08 mmol, 12%) of **A4D3** as a yellow solid.

MACROCYCLE A4D3



Compound 52b. To a 200-mL (24/40) round-bottomed flask equipped with a magnetic stir bar was added **51** (1.65 g, 10 mmol), 2.1 mL (12 mmol) of DIPEA, and 55 mL of MeCN. While stirring at room temperature, *tert*-butyl (4-aminobutyl)carbamate (1.88 g, 10 mmol) was added and the solution was stirred for 12 h. The solution was initially colorless and white precipitate began to form after a few hours. The precipitate was filtered off and rinsed with water (30 mL) and dried under nitrogen to afford 2.59 g (8.2 mmol, 82%) of **52b** as a white solid.

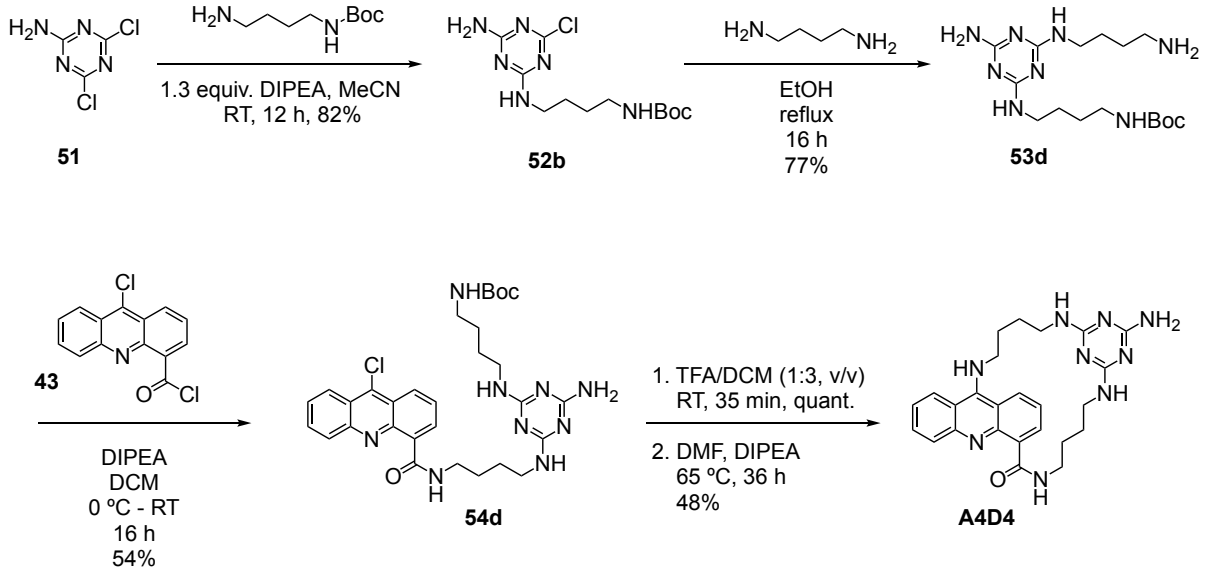
Compound 53b. The title compound was prepared in similar conditions as described for compound **53a** from 5.28 g (17.1 mmol) of **52b** and 14.3 mL (171 mmol) of 1,3-diaminopropane, affording 3.39 g (9.57 mmol, 56%) of **59** as an off-white solid.

Compound 54b. The title compound was prepared in similar conditions as described for compound **54a** from 1.03 g (4.31 mmol) of **43** and 0.995 g (2.81 mmol) of **53b**, affording 0.564 g (0.95 mmol, 34%) of **54b** as a dark yellow solid.

Macrocycle A4D3. The title compound was prepared in similar as described for macrocycle **A3D3** from 0.301 g (0.51 mmol) of **54b**, affording 0.178 g (0.34 mmol, 66%) of **A4D3** as a

yellow solid.

MACROCYCLE A4D4

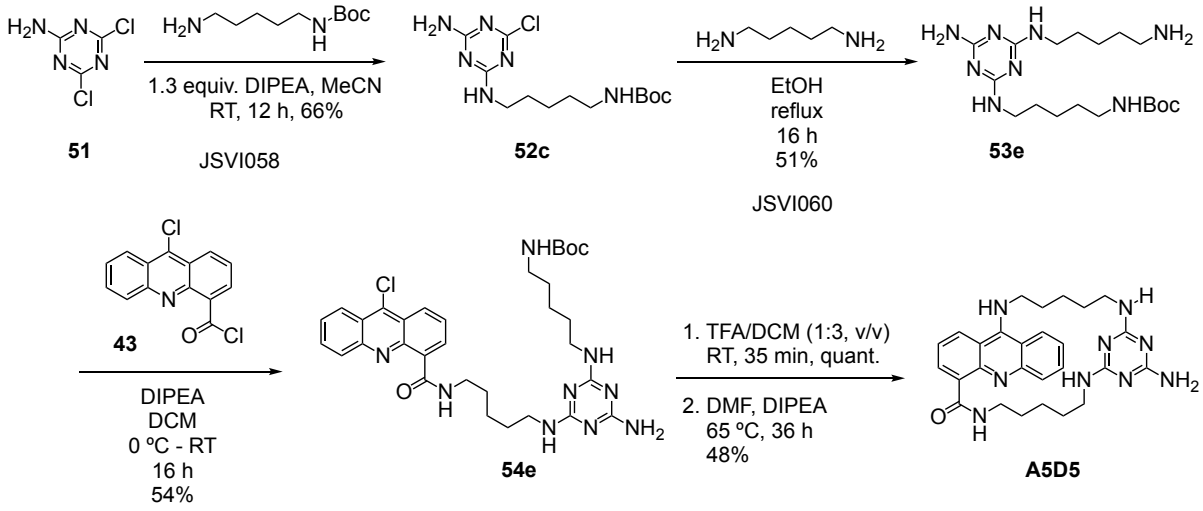


Compound 53d. The title compound was prepared in similar conditions as described for compound **53a** from 3.48 g (11 mmol) of **52b** and 11 mL (110 mmol) of 1,4-diaminobutane, affording 3.07 g (8.33 mmol, 77%) of **53d** as a white solid.

Compound 54d. The title compound was prepared in similar conditions as described for compound **54a** from 4.71 mmol of **43** and 1.23 g (3.36 mmol) of **53d**, affording 0.564 g (0.95 mmol, 34%) of **54d** as a dark yellow solid.

Macrocycle A4D4. The title compound was prepared in similar as described for macrocycle **A3D3** from **6** 0.343 g (0.56 mmol) of **60**, affording 0.146 g (0.27 mmol, 48%) of **A4D4** as a yellow solid.

MACROCYCLE A5D5



Compound 52c. The title compound was prepared in similar conditions as described for compound **52a** from 1.94 g (11.8 mmol) of **51** and 2.05 g (11.8 mmol) of *tert*-butyl (5-aminopentyl)carbamate, affording 3.12 g (10.1 mmol, 86%) of **52c** as a white solid.

Compound 53e. The title compound was prepared in similar conditions as described for compound **53a** from 3.10 g (10.2 mmol) of **52c** and 8.5 mL (102.4 mmol) of 1,5-diaminopentane, affording 1.91 g (5.61 mmol, 55%) of **53e** as an off-white solid.

Compound 54e. The title compound was prepared in similar conditions as described for compound **54a** from 1.12 g (4.68 mmol) of **43** and 1.00 g (2.94 mmol) of **53e**, affording 1.40 g (2.41 mmol, 82%) of **54e** as a dark yellow solid.

Macrocycle A5D5. The title compound was prepared in similar as described for macrocycle **A3D3** from 0.387 g (0.67 mmol) of **54e**, affording 0.16 g (0.31 mmol, 47%) of **A5D5** as a dark yellow solid.

4.3. Appendix

4.3.1. NMR Spectra

Figure 31. ^{1}H NMR spectrum of ligand **11**.

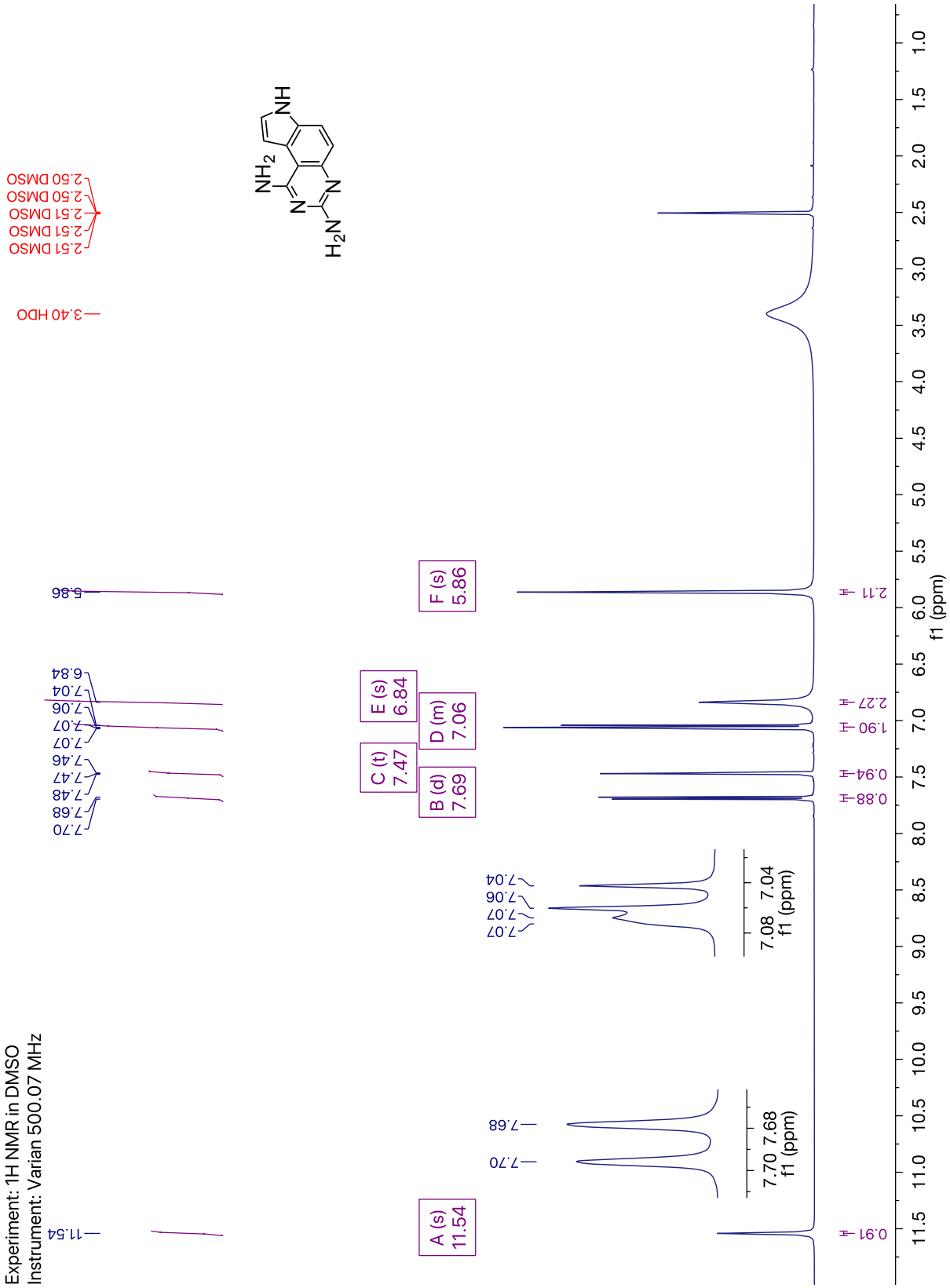


Figure 32. ^{1}H NMR spectrum of ligand **39**.

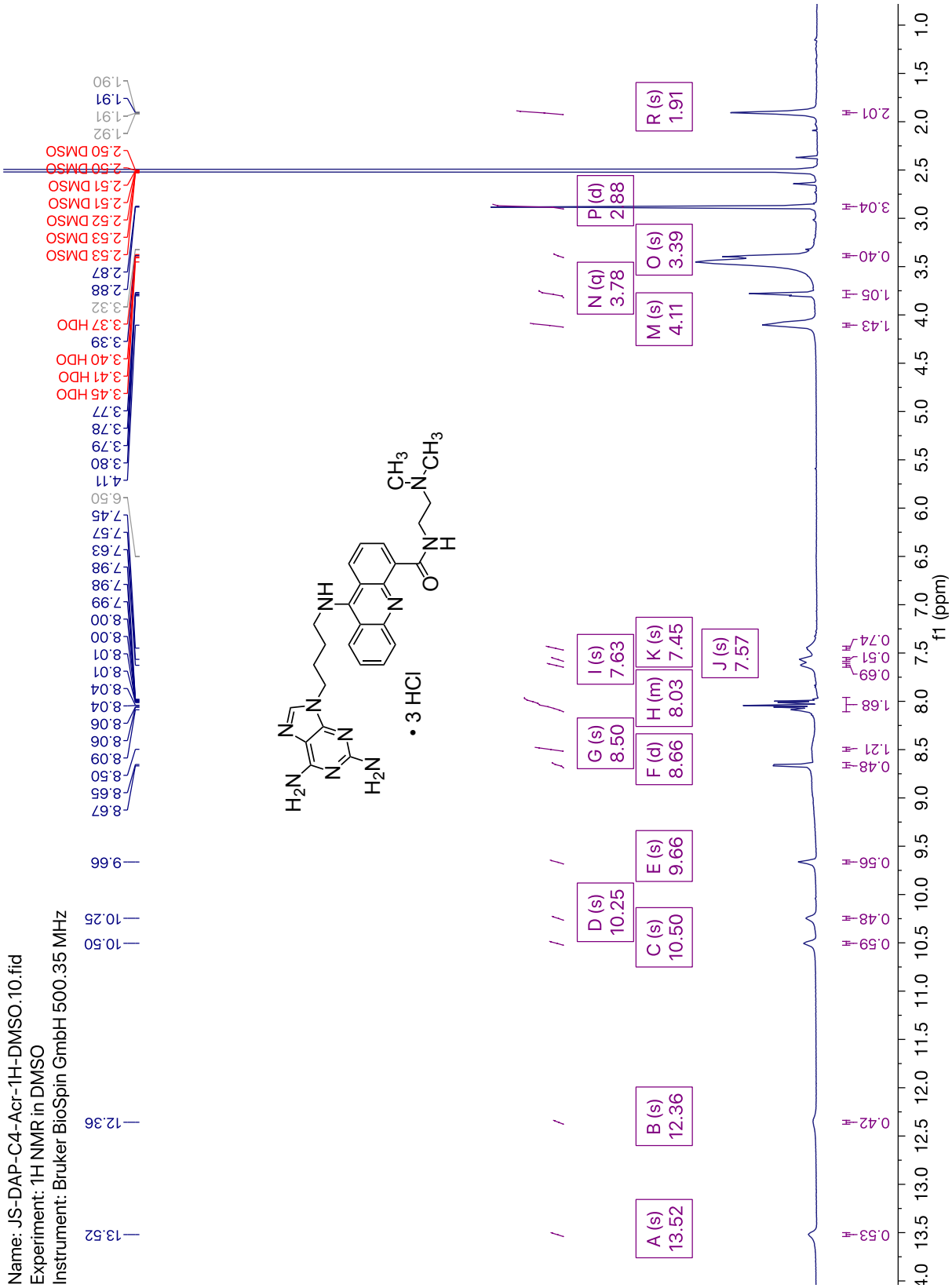


Figure 33. ^{1}H NMR spectrum of ligand **47**.

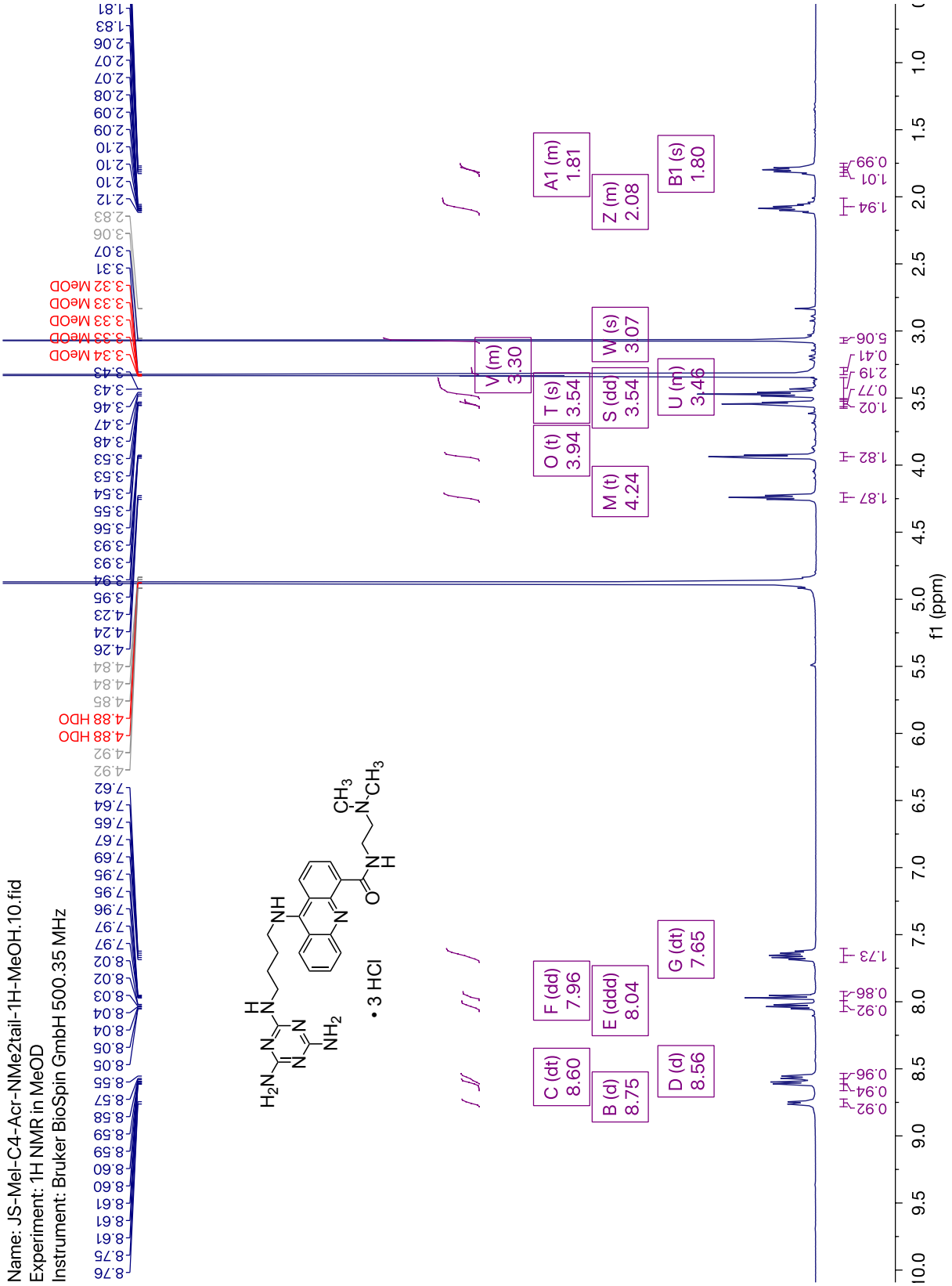


Figure 34. ^{1}H NMR spectrum of ligand **56**.

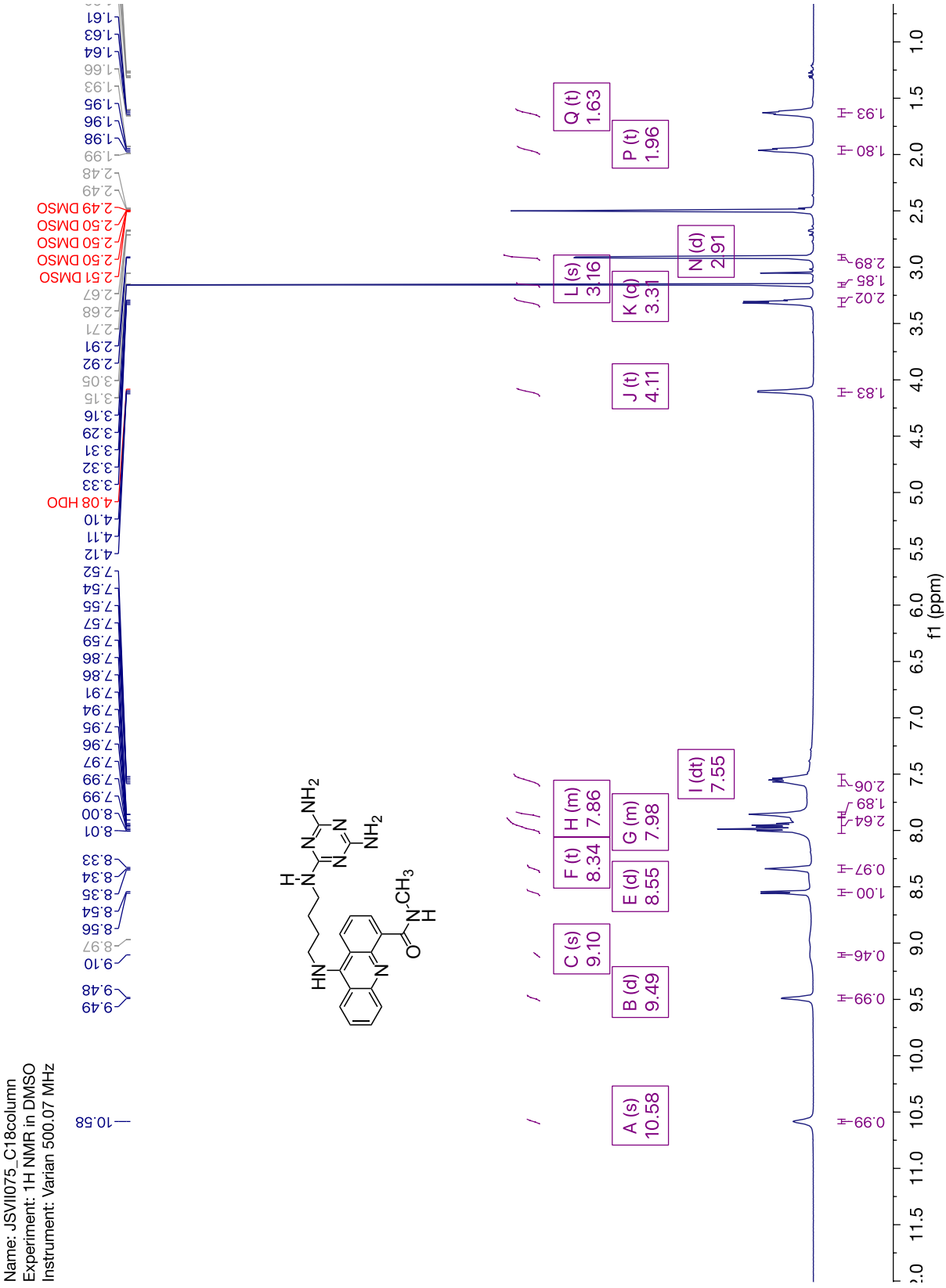


Figure 35. [1]H NMR spectrum of ligand **A3D3**.

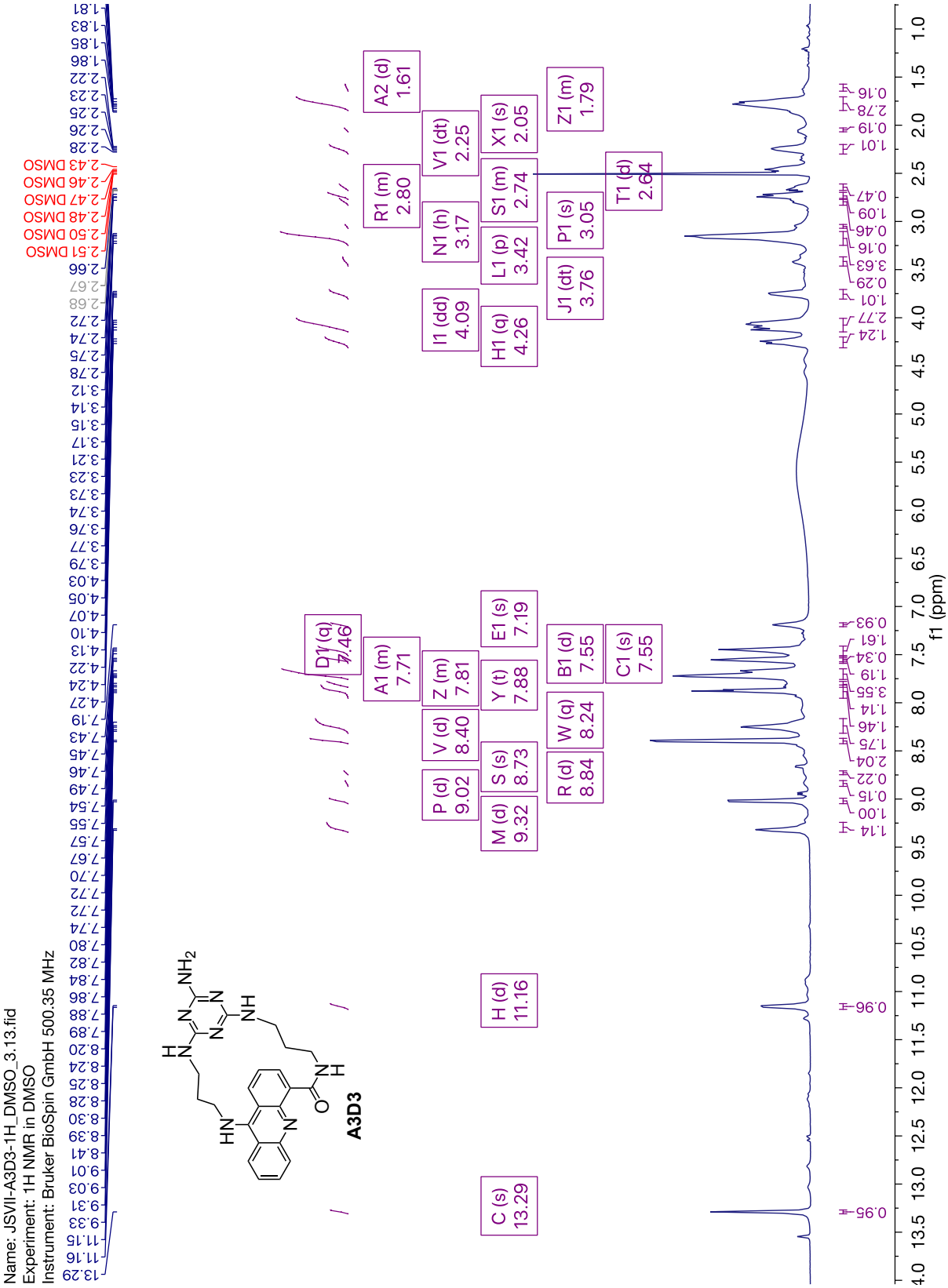


Figure 36. ^{1}H COSY NMR spectrum of ligand **A3D3**.

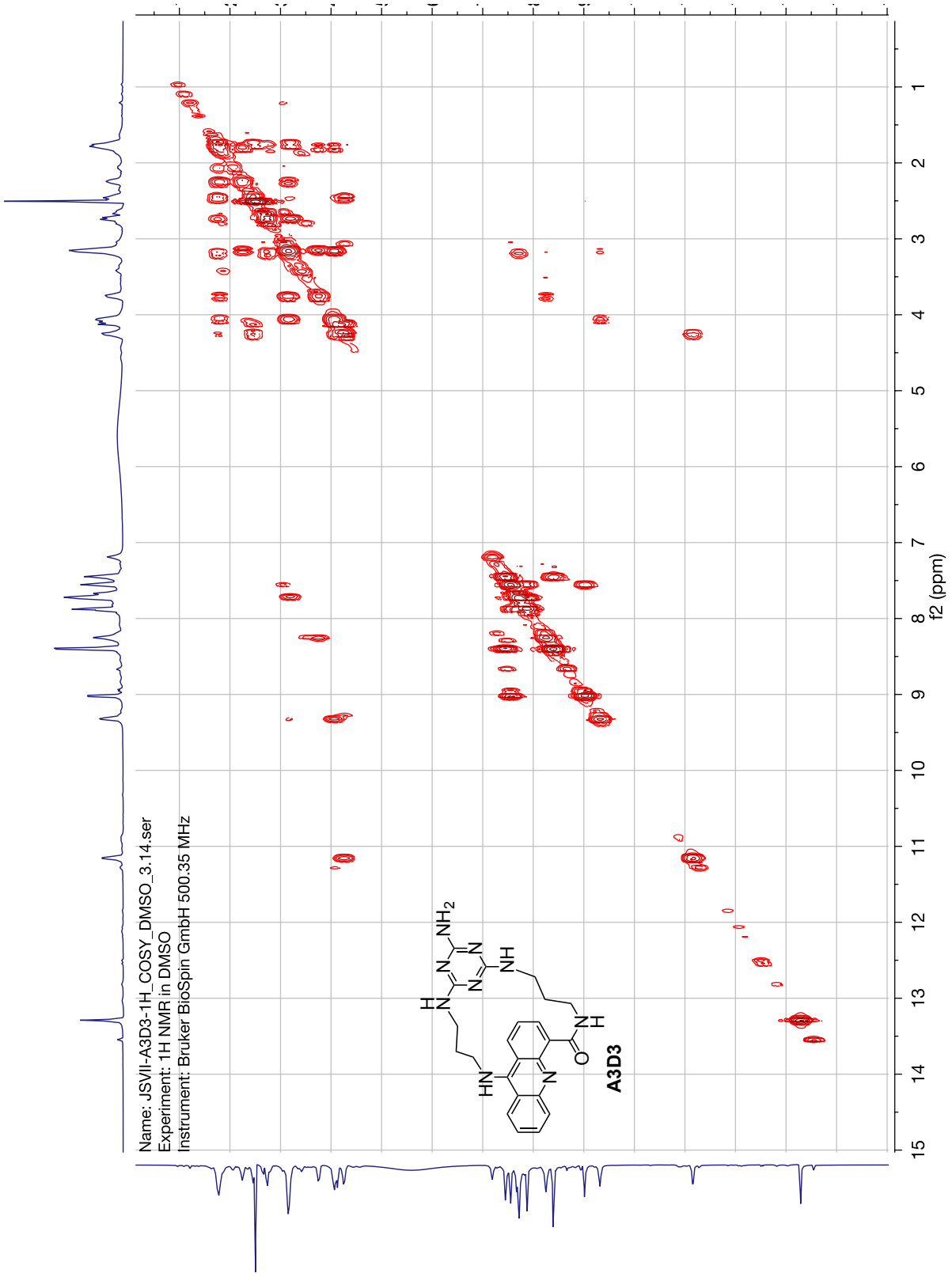


Figure 37. ^{13}C NMR spectrum of ligand **A3D3**.

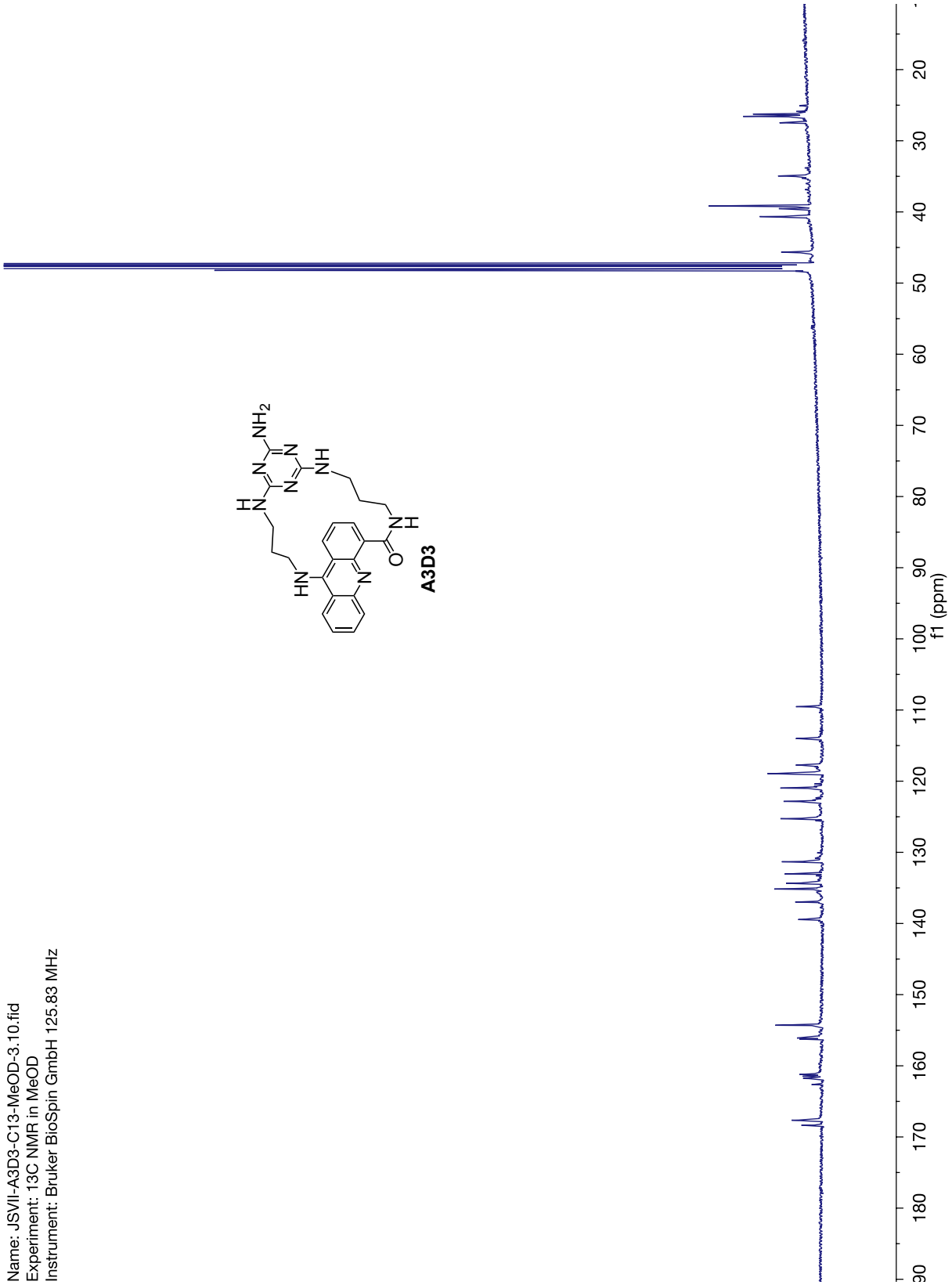


Figure 38. [1]H NMR spectrum of ligand **A3D4**.

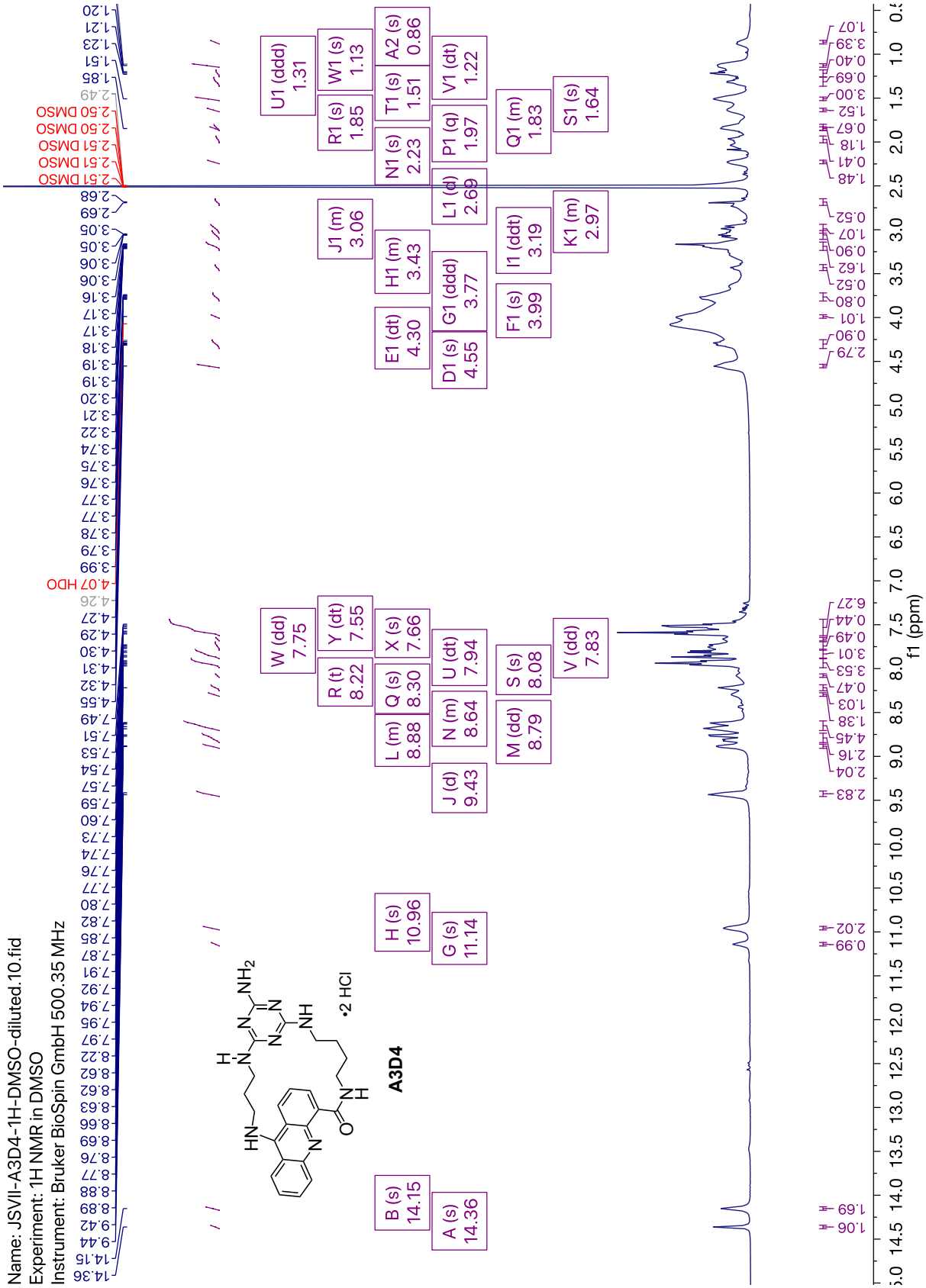


Figure 39. ^{1}H COSY NMR spectrum of ligand **A3D4**.

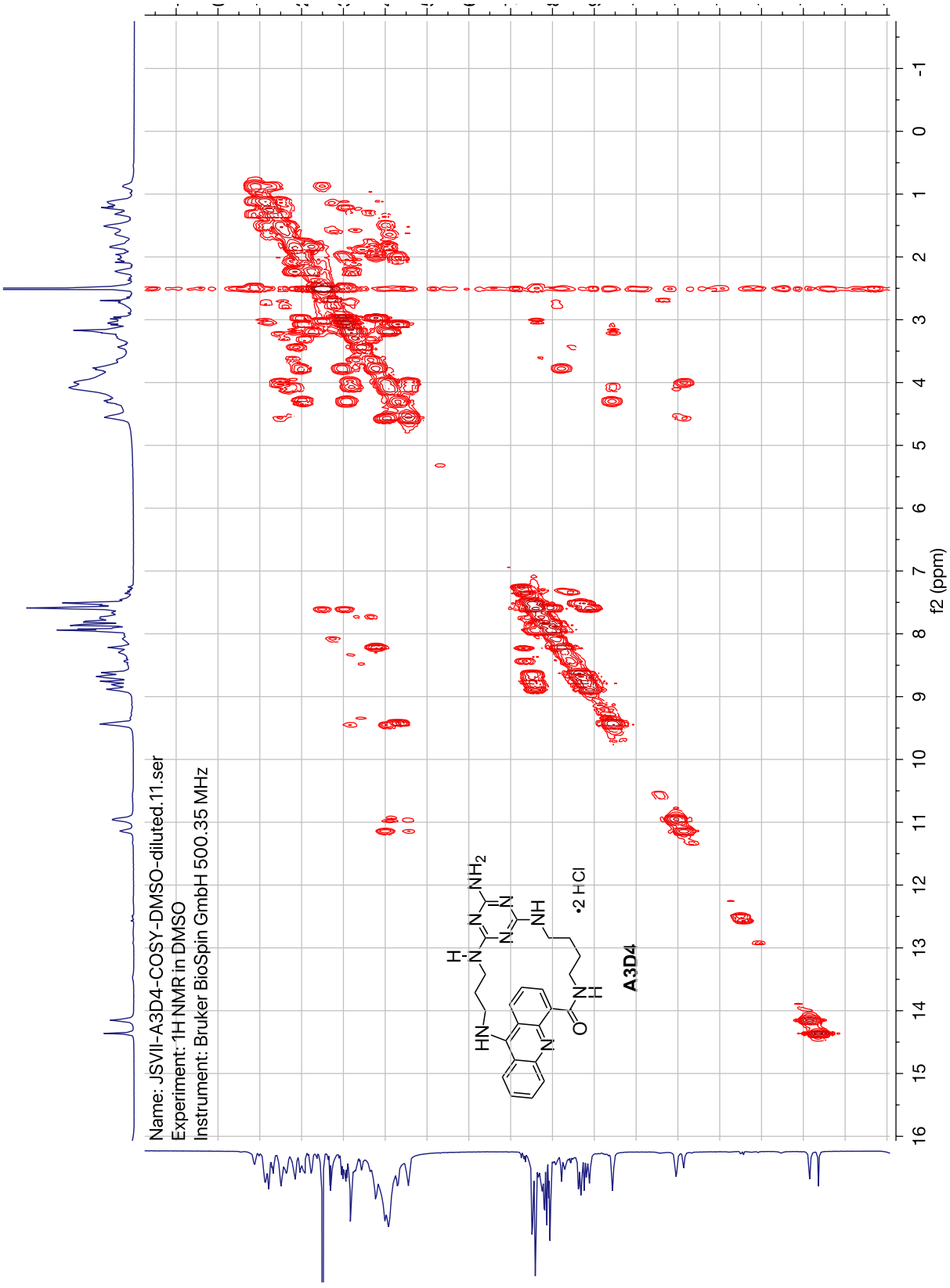


Figure 40. ^{13}C NMR spectrum of ligand **A3D4**.

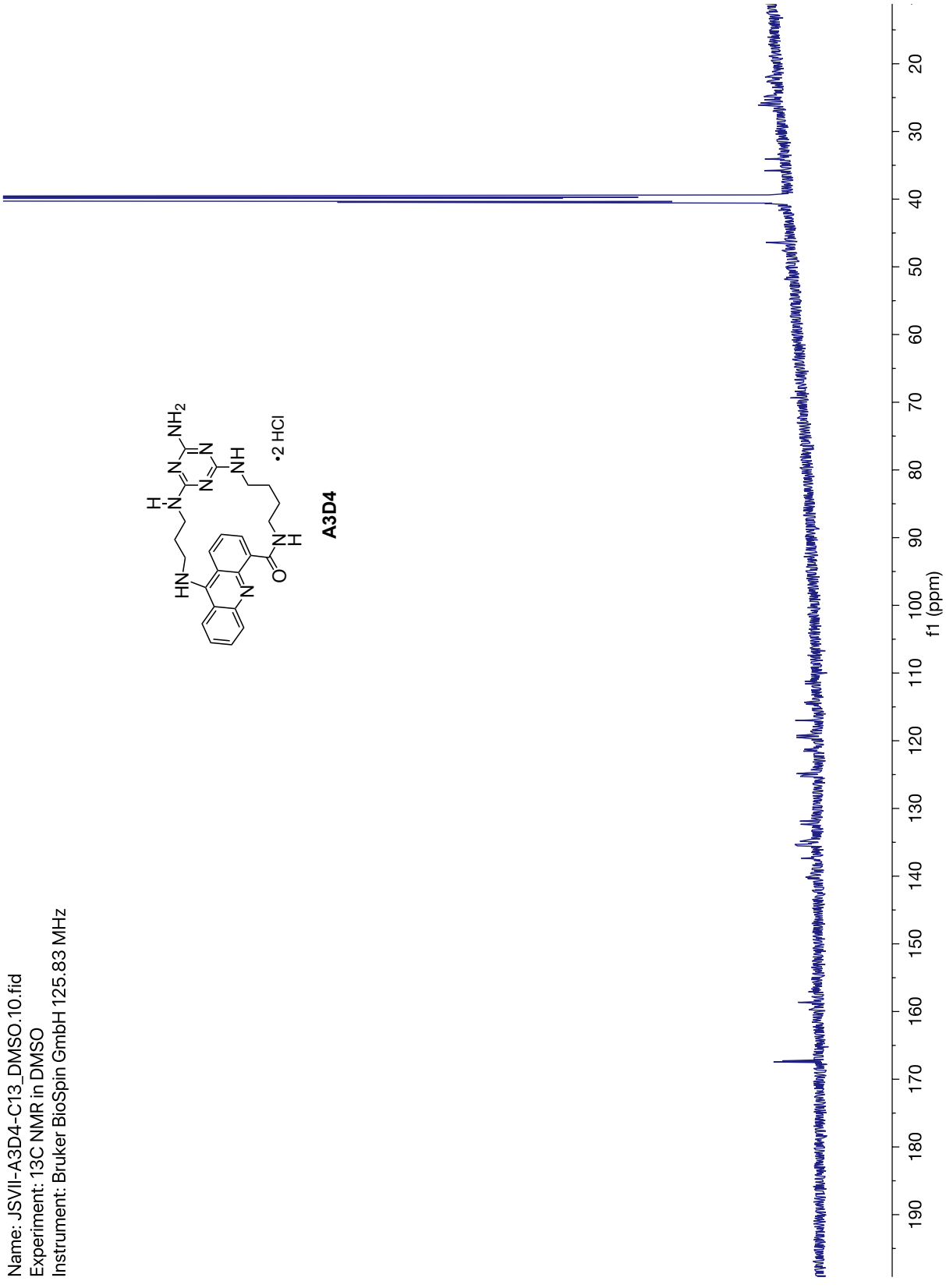


Figure 41. [1]H NMR spectrum of ligand **A4D3**.

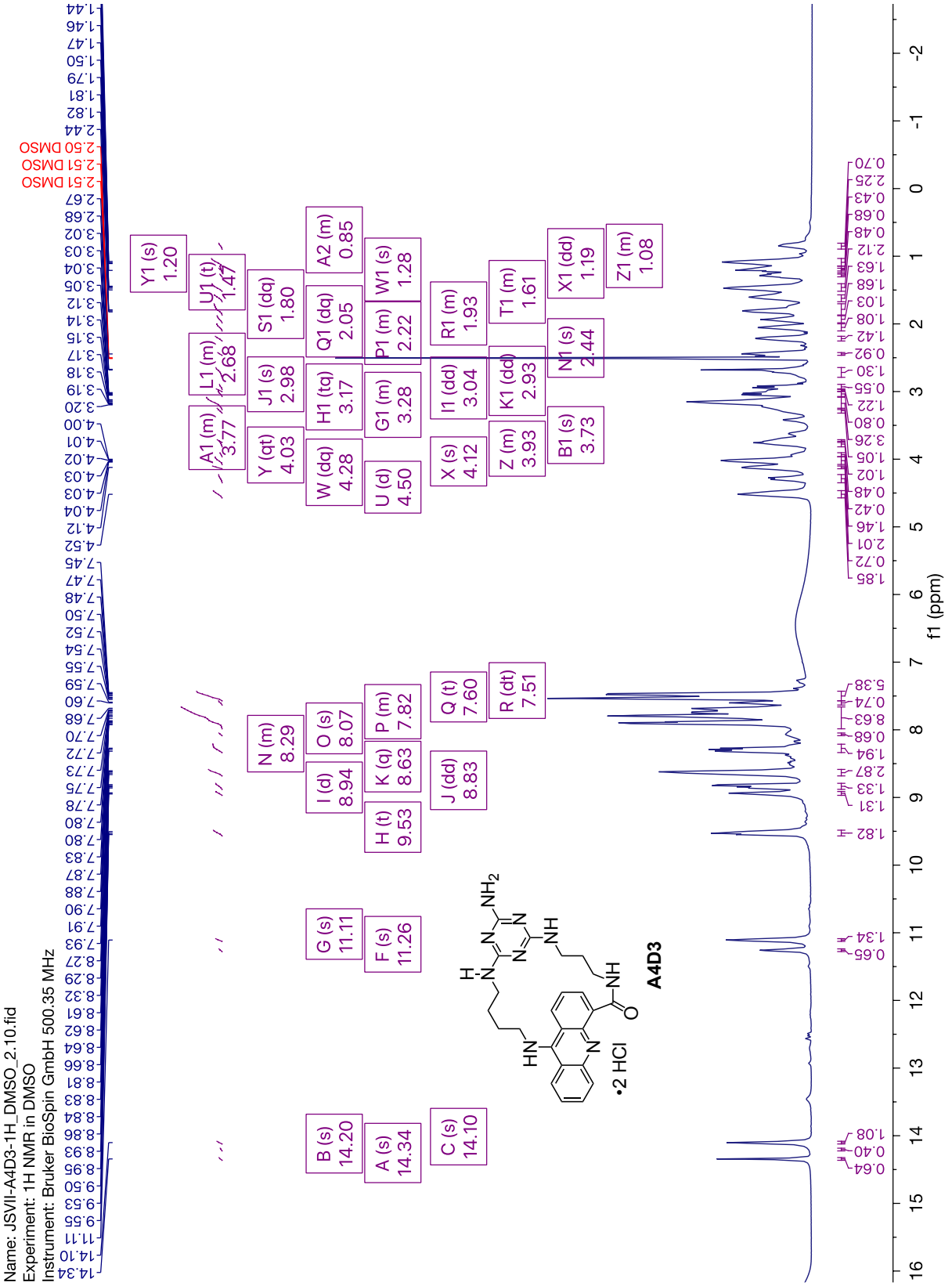


Figure 42. [1]H COSY NMR spectrum of ligand **A4D3**.

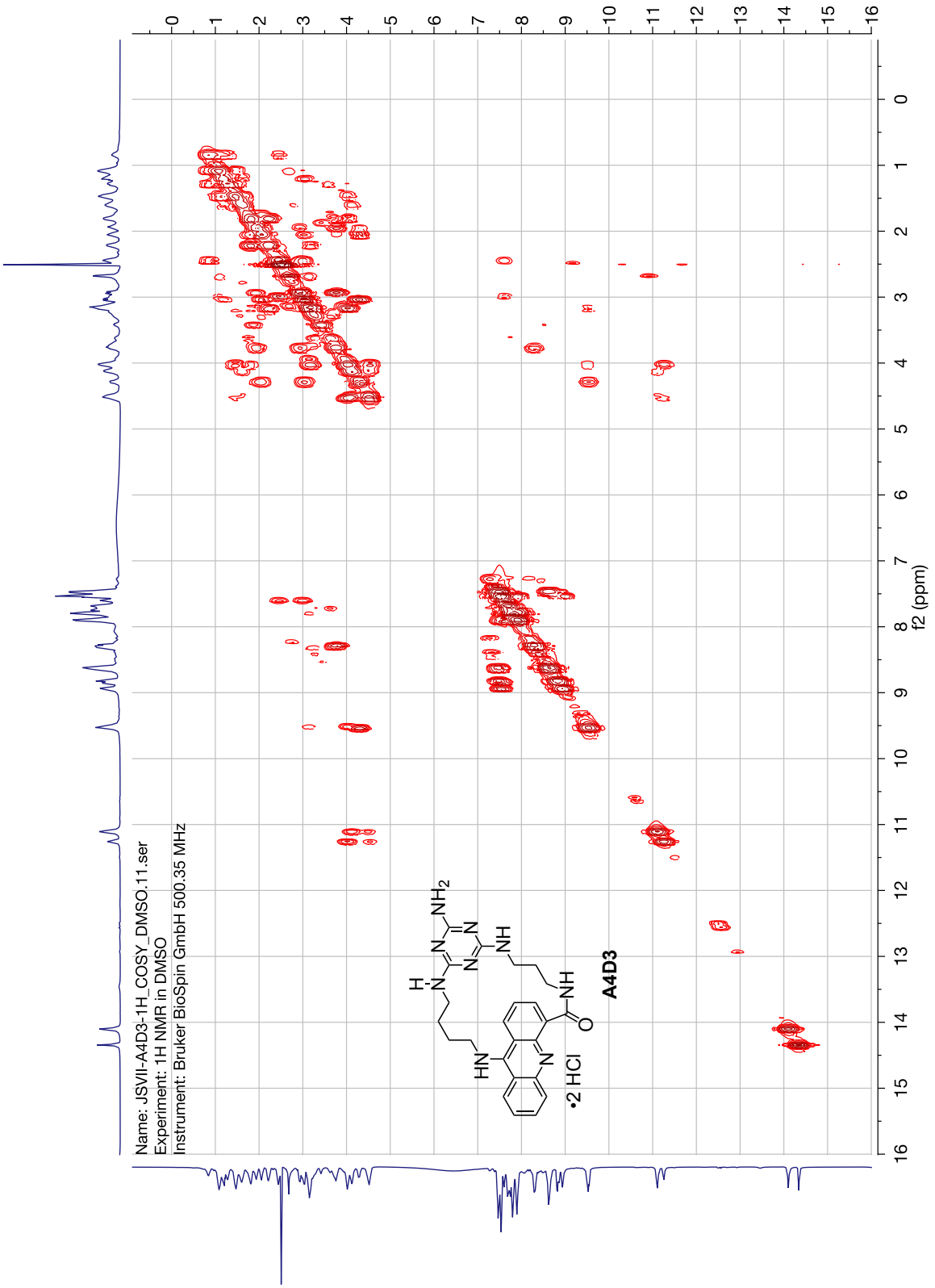


Figure 43. ^{13}C NMR spectrum of ligand **A4D3**.

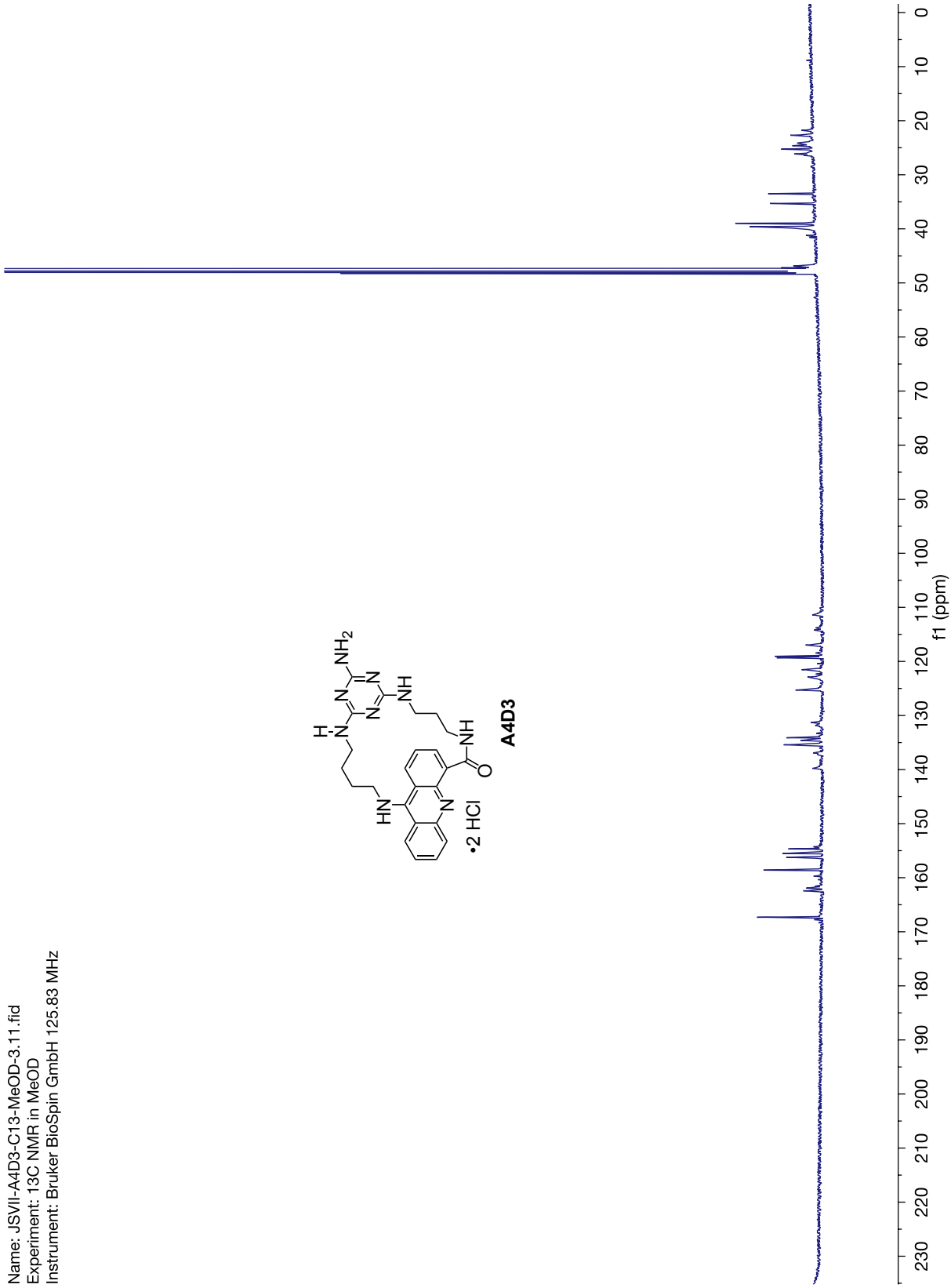


Figure 44. ^{1}H NMR spectrum of ligand **A4D4**.

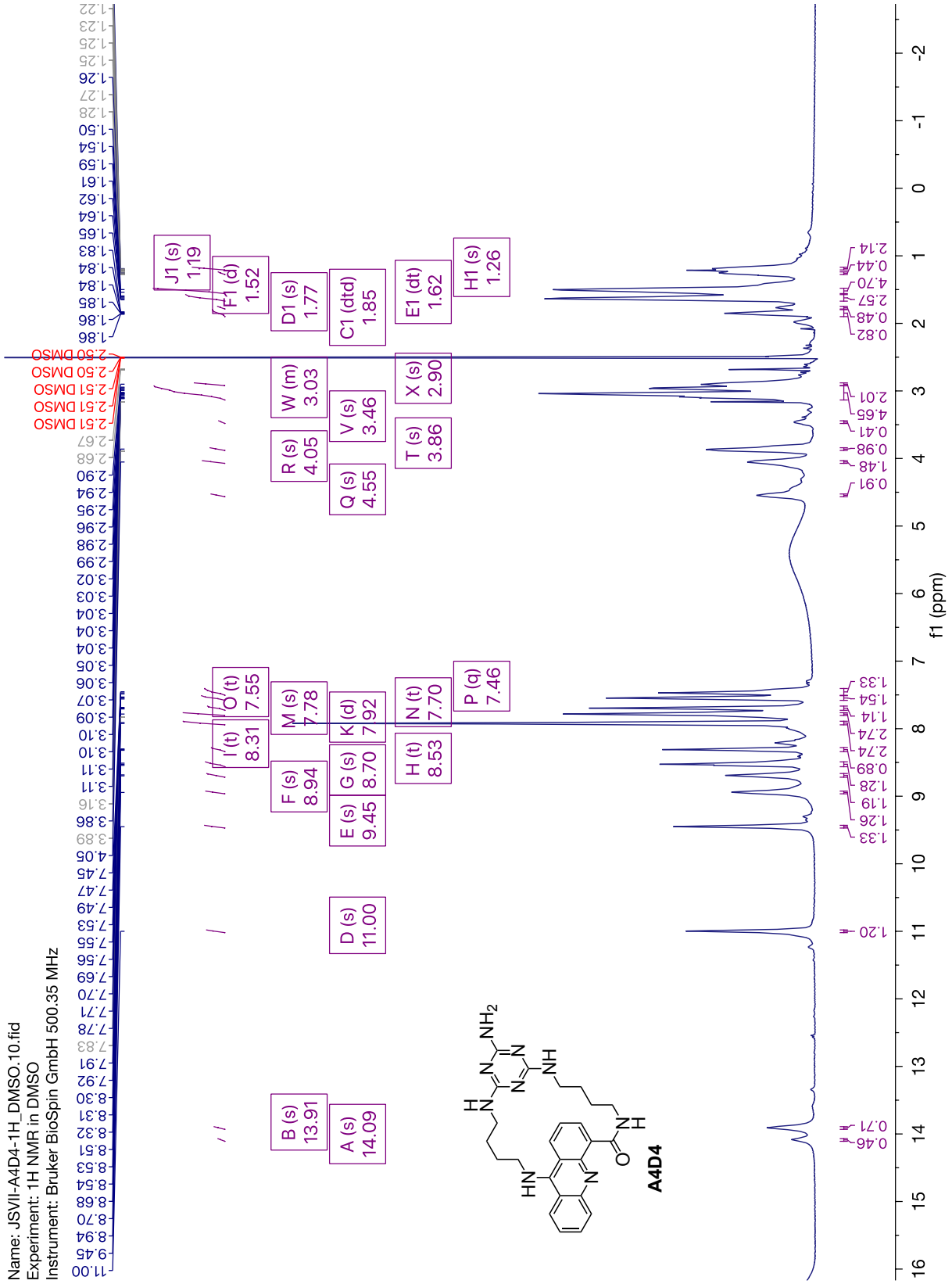


Figure 45. [1]H COSY NMR spectrum of ligand **A4D4**.

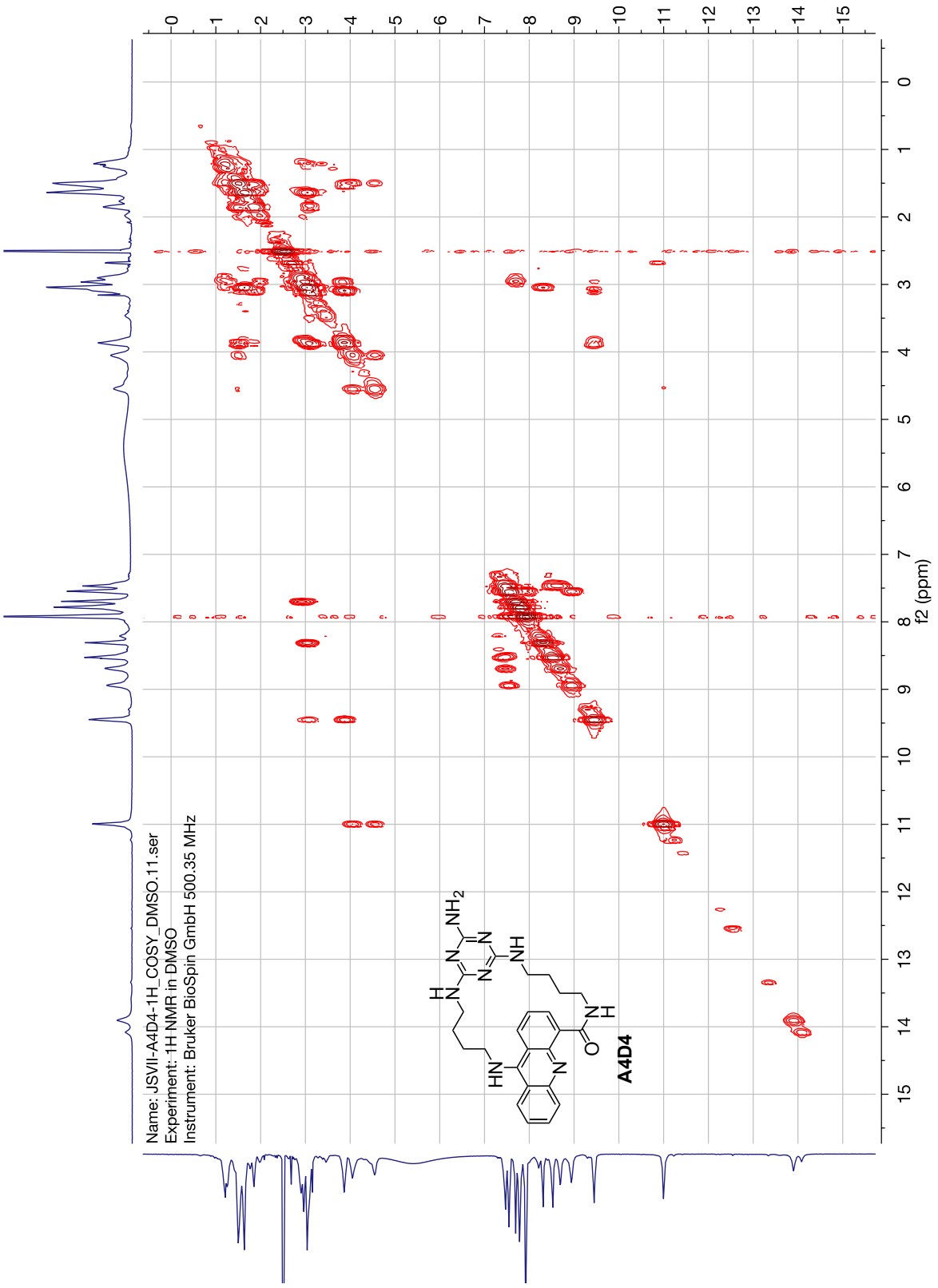


Figure 46. ^{13}C NMR spectrum of ligand **A4D4**.

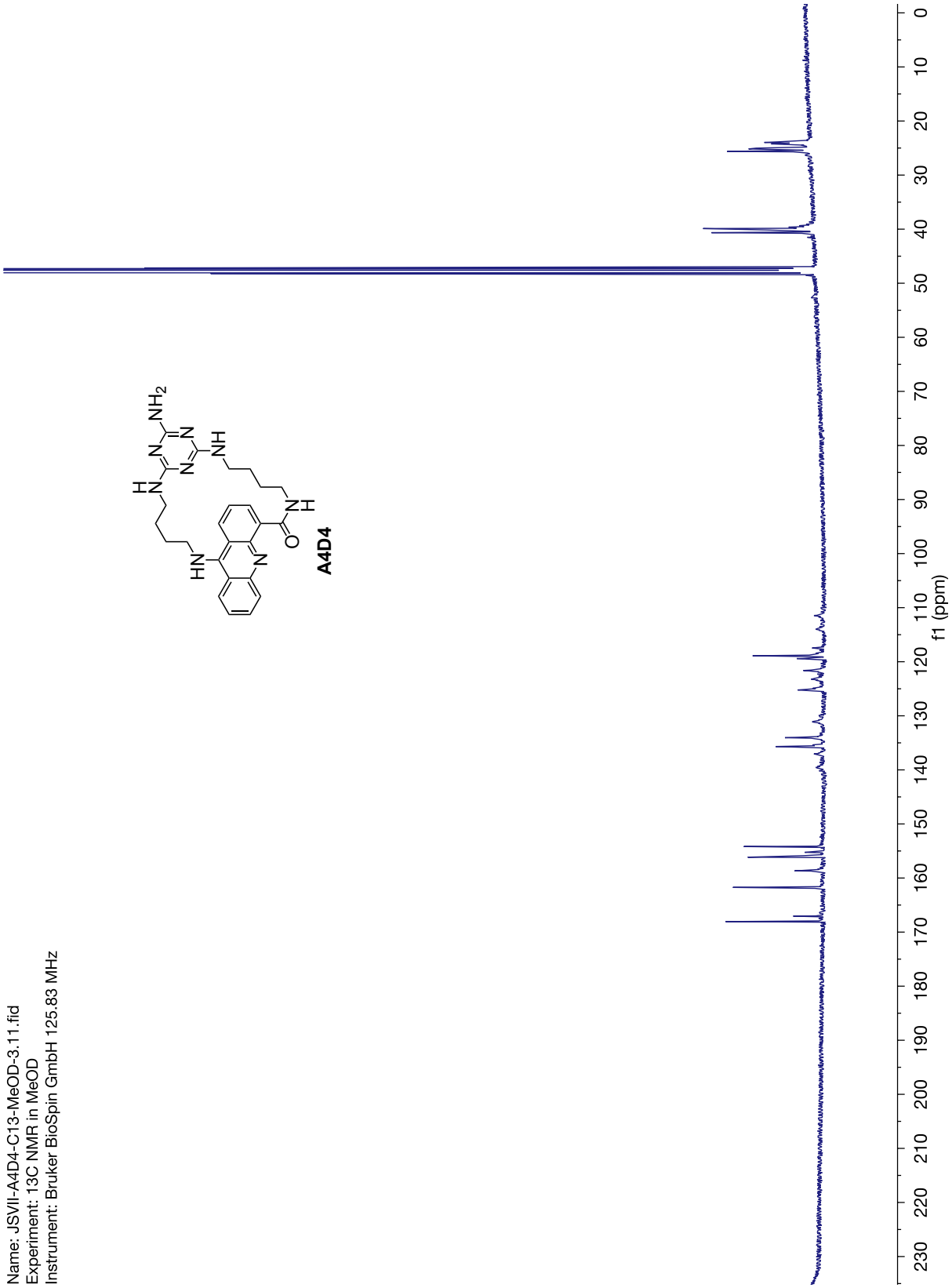


Figure 47. [1]H COSY NMR spectrum of ligand **A5D5**.

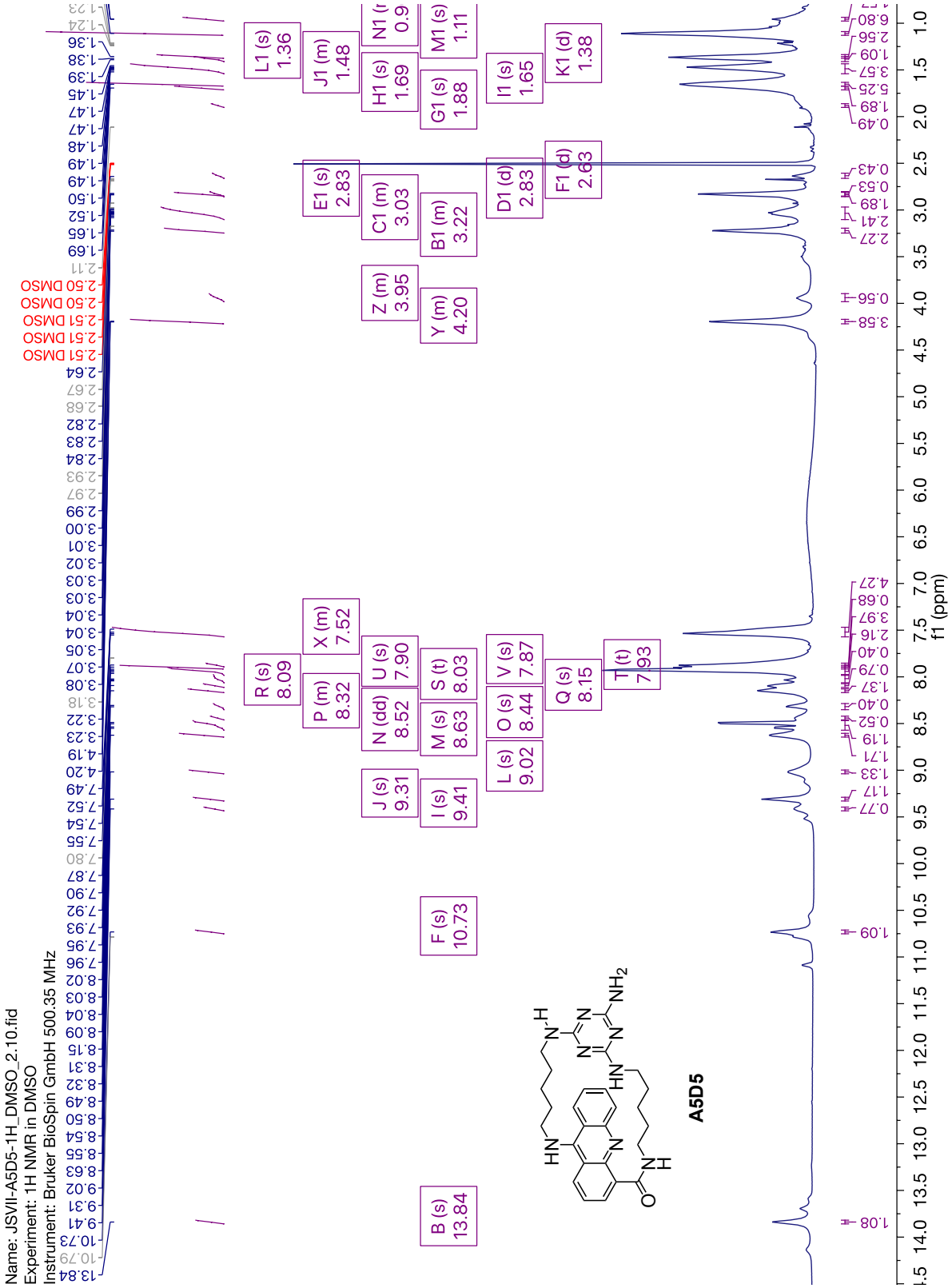


Figure 48. ^{1}H COSY NMR spectrum of ligand **A5D5**.

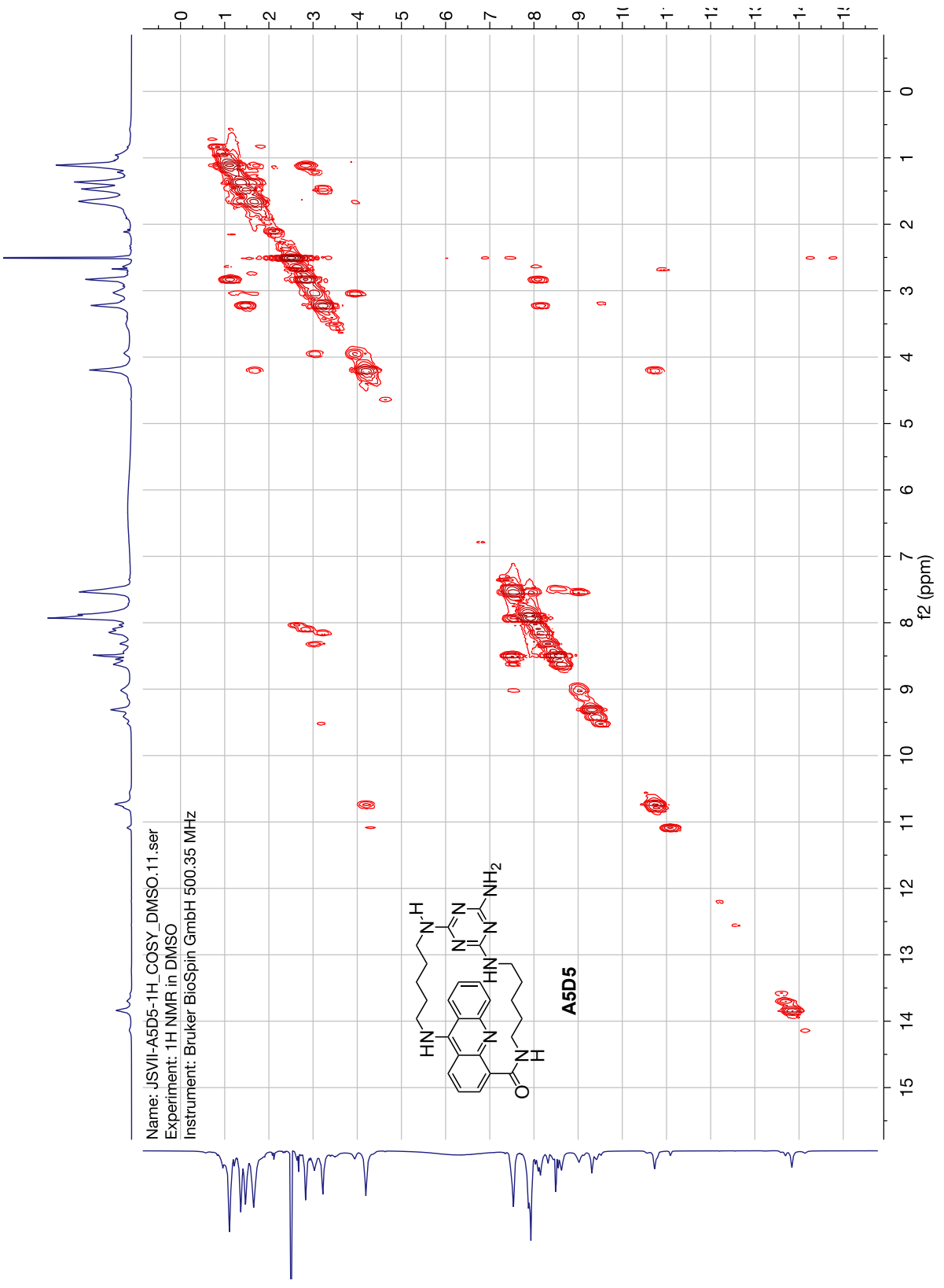


Figure 49. ^{13}C NMR spectrum of ligand **A5D5**.

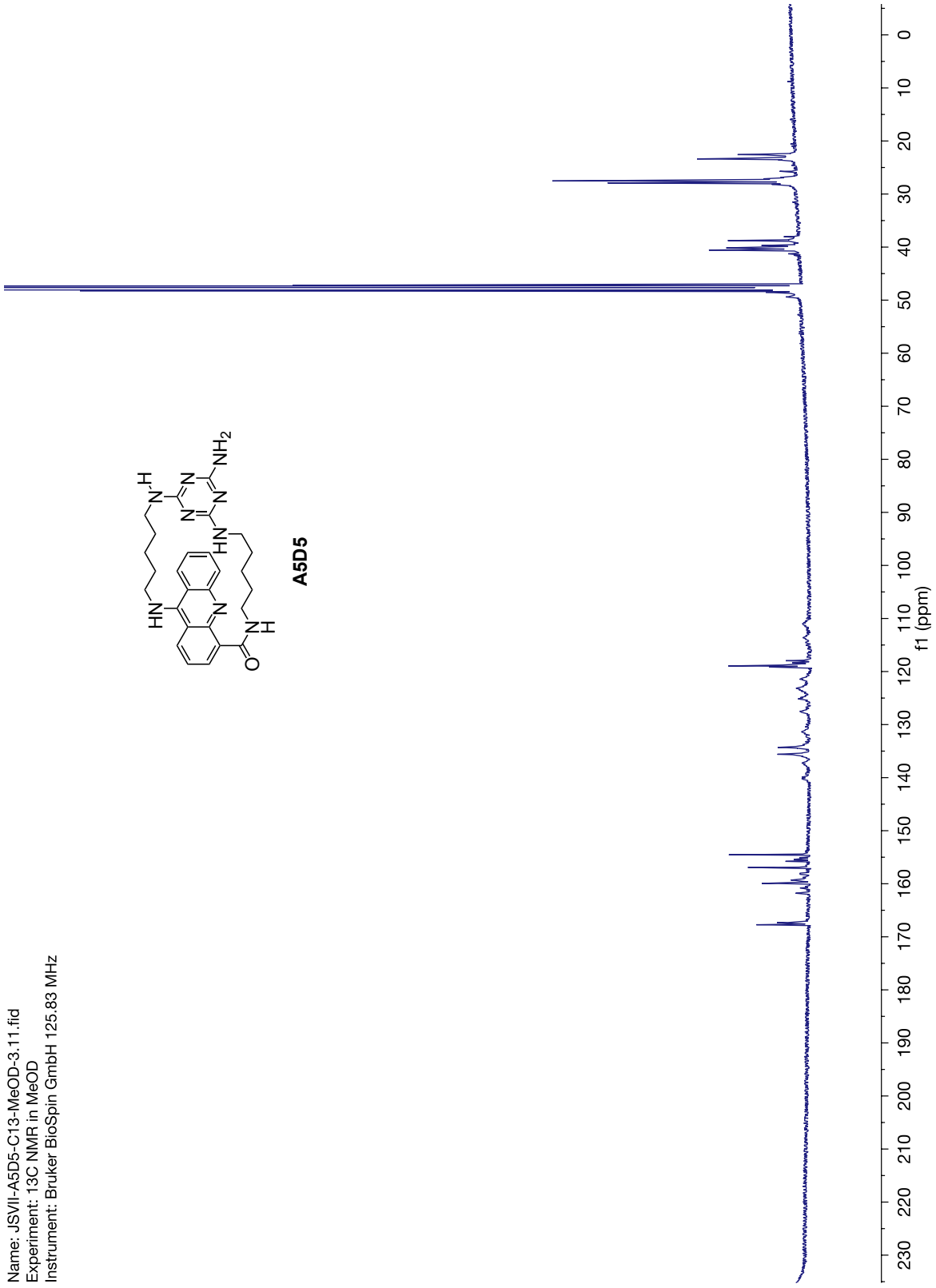


Figure 50. ITC isotherms of the macrocyclic ligands with dsDNA.

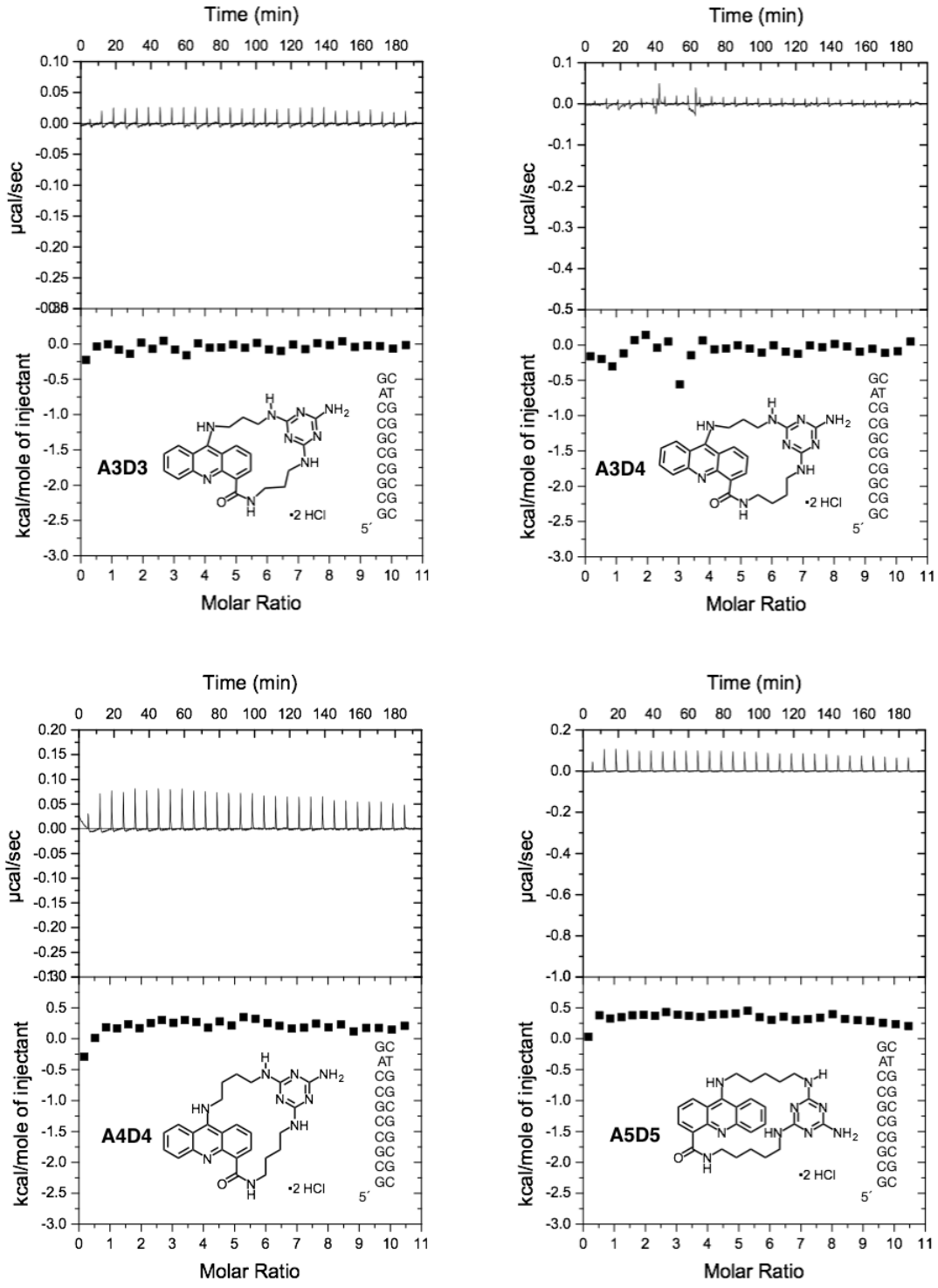


Figure 50. (cont.)

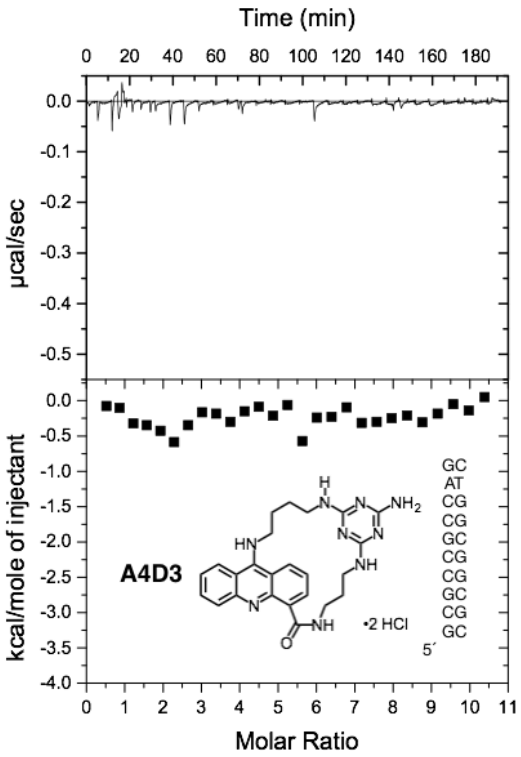


Figure 51. ITC isotherms of the macrocyclic ligands with $d(CTG)_2$.

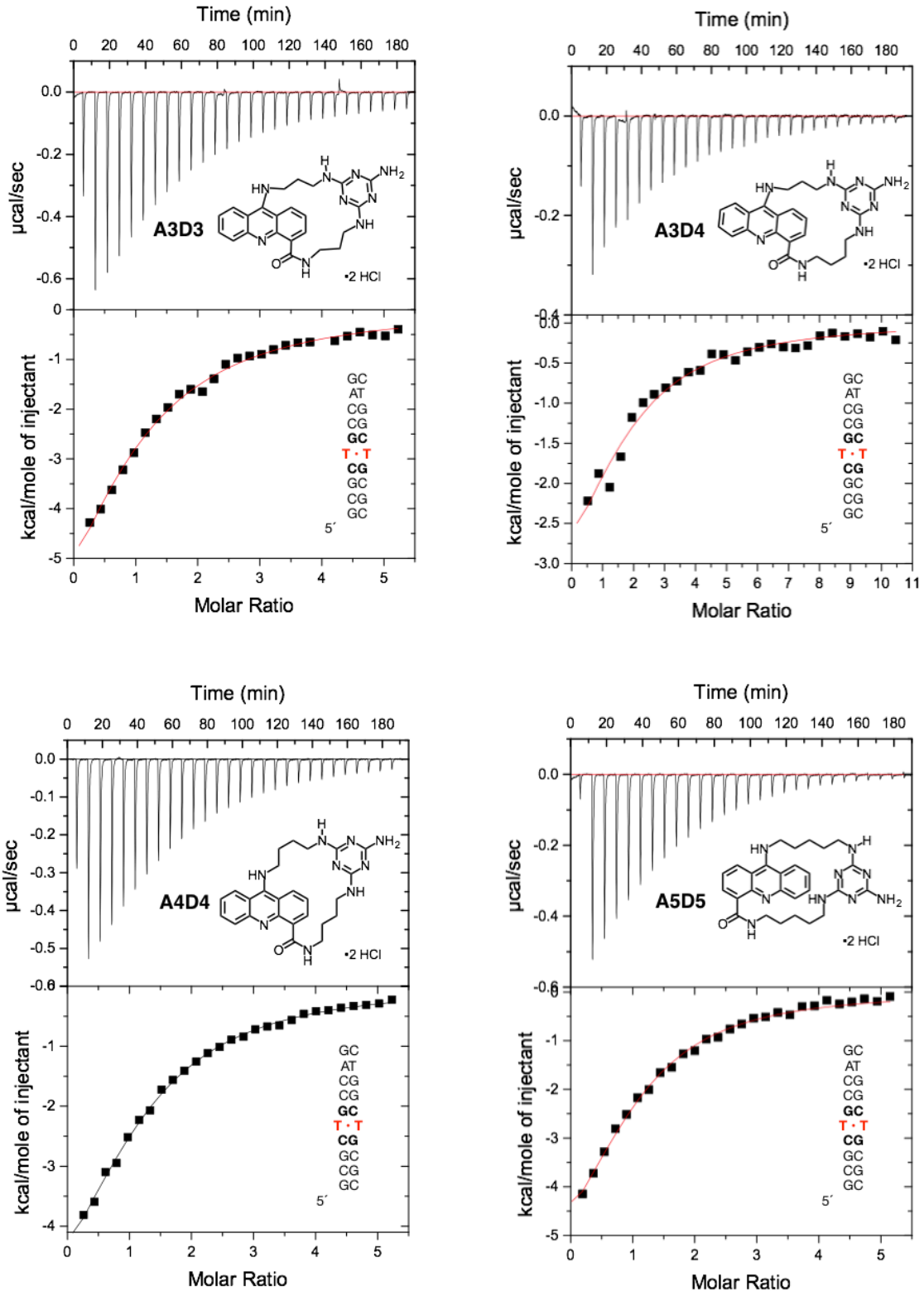


Figure 51. (cont.)

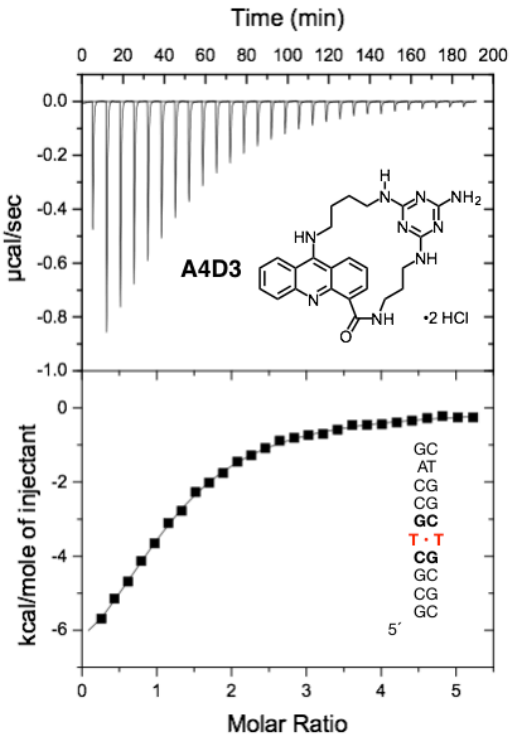


Figure 52. ITC isotherms of the macrocyclic ligands with d(CCG)$_2$.

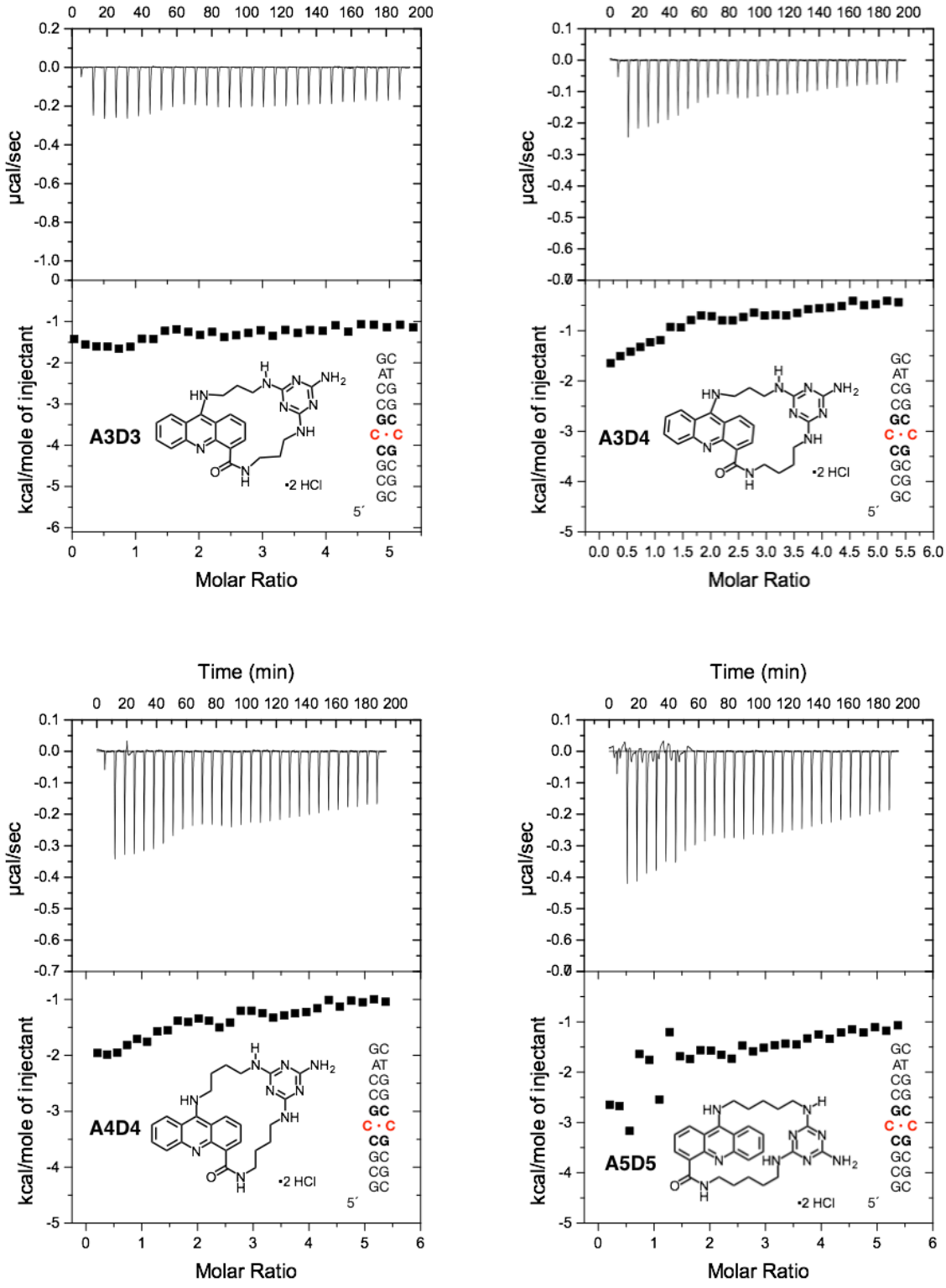


Figure 52. (cont.)

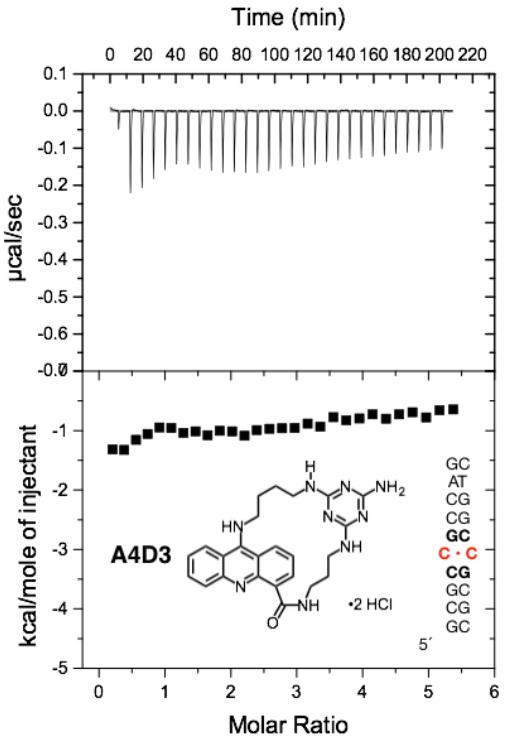


Figure 53. ITC isotherms of the macrocyclic ligands with r(CUG)$_2$.

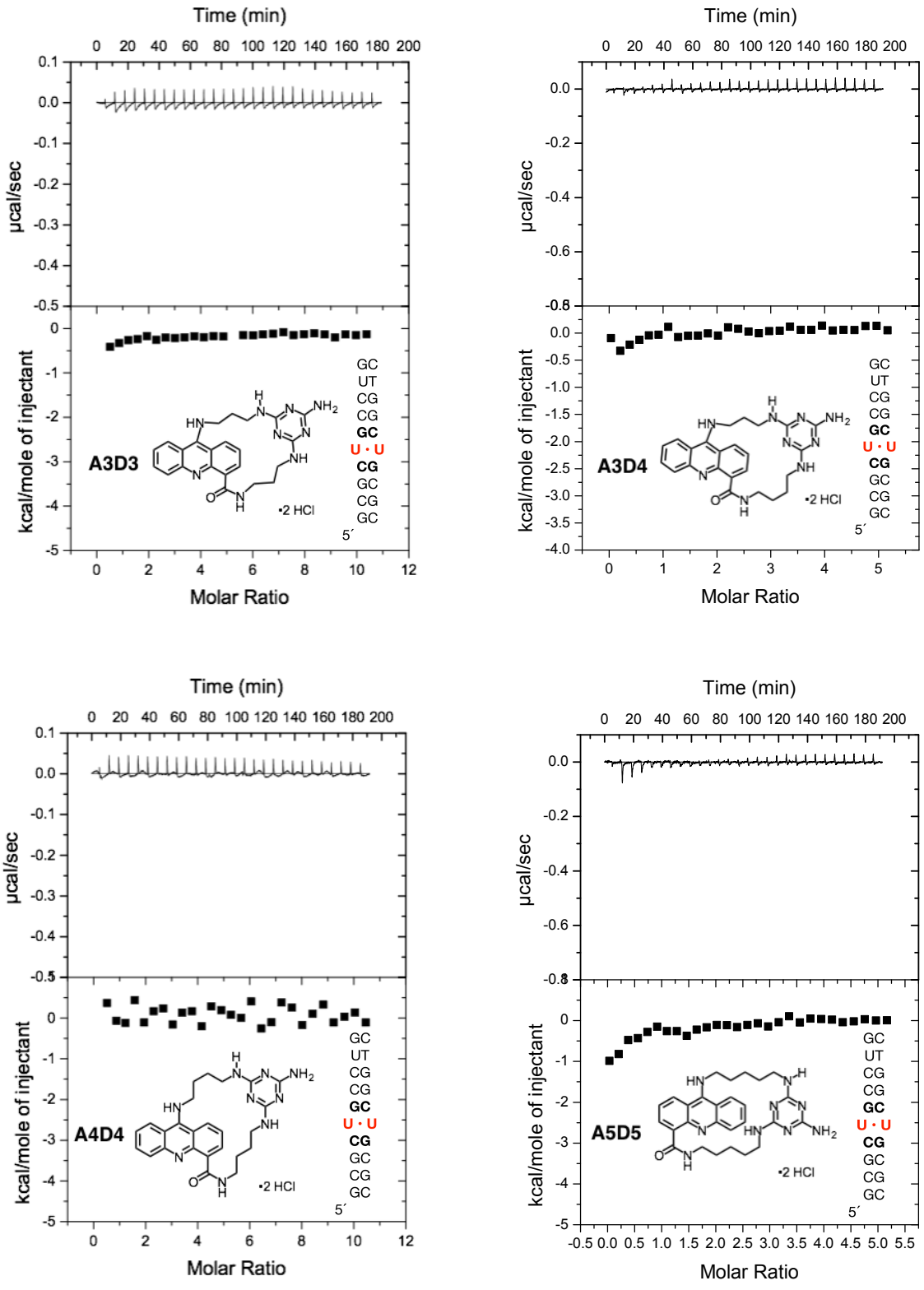


Figure 53. (cont.)

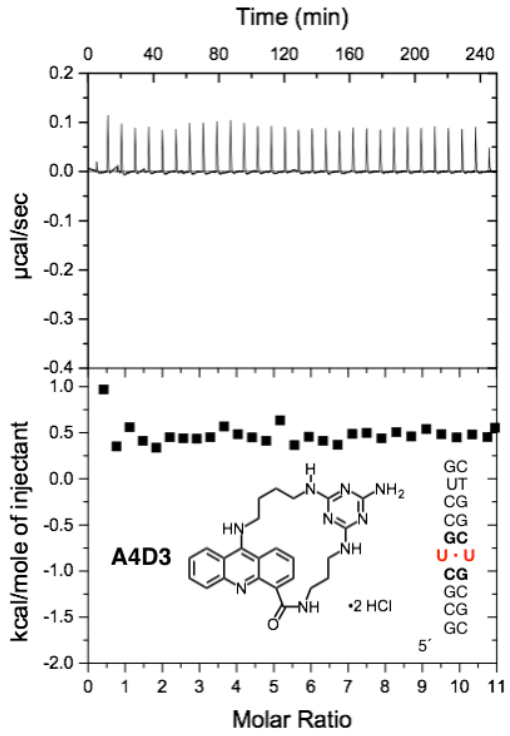


Figure 54. ITC isotherms of the macrocyclic ligands with d(CAG)$_2$.

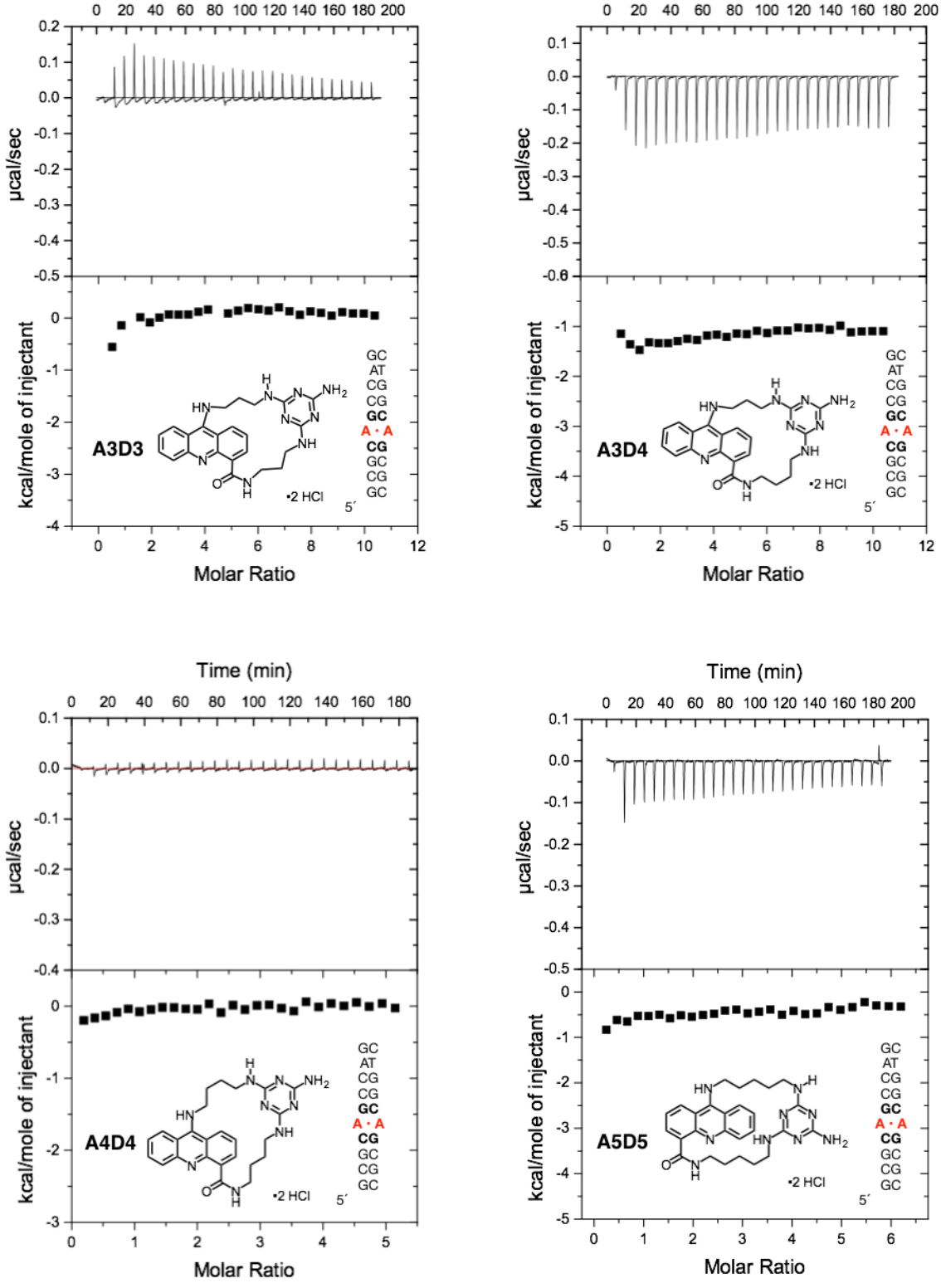
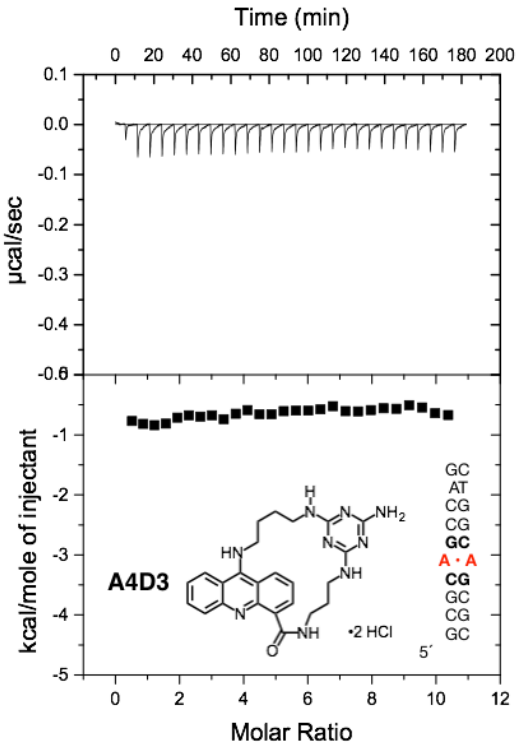


Figure 54. (cont.)



REFERENCES

(1) Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18.

(2) Tsai, C.-C.; Jain, S. C.; Sobell, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72* (2), 628.

(3) Shieh, H.-S.; Berman, H. M.; Dabrow, M.; Neidle, S. *Nucleic Acids Research* **1980**, *8* (1), 85.

(4) Hollstein, U. *Chem. Rev.* **1974**, *74* (6), 625.

(5) Waring, M. J.; Wakelin, L. P. G. *Nature* **1974**, *252* (5485), 653.

(6) Arcamone, F.; Penco, S.; Orezzi, P.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203* (494), 1064.

(7) *Small Molecule DNA and RNA Binders*; Demeunynck, M., Bailly, C., Wilson, W. D., Eds.; Wiley-VCH: Weinheim, 2003.

(8) Paul, A.; Bhattacharya, S. *Current Science* **2012**.

(9) Xie, Y.; Tam, V. K.; Tor, Y. In *The Chemical Biology of Nucleic Acids*; Mayer, G., Ed.; The Chemical Biology of Nucleic Acids; John Wiley & Sons, Ltd: Chichester, UK, 2010; Vol. 171, pp 115–140.

(10) Escudé, C.; Garestier, T.; Sun, J.-S. *Drug interaction with triple-helical nucleic acids*; Methods in Enzymology; Elsevier, 2001; Vol. 340, pp 340–357.

(11) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (18), 11593.

(12) Saenger, W. In *Principles of Nucleic Acid Structure*; Springer Advanced Texts in Chemistry; Springer New York: New York, NY, 1984; pp 350–367.

(13) Crothers, D. M. *Biopolymers* **1968**, *6* (4), 575.

(14) Dervan, P. B. *Bioorganic & Medicinal Chemistry* **2001**, *9* (9), 2215.

(15) Dervan, P. B. *Science* **1986**, *232* (4749), 464.

(16) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 284.

(17) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116* (18), 7983.

(18) *Chemical Biology of Nucleic Acids*; Erdmann, V. A., Markiewicz, W. T., Barciszewski, J., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014.

(19) Granzhan, A.; Kotera, N.; Teulade-Fichou, M.-P. *Chem. Soc. Rev.* **2014**, *43* (10), 3630.

(20) De Bont, R.; van Larebeke, N. *Mutagenesis* **2004**, *19* (3), 169.

(21) Kunkel, T. A.; Erie, D. A. *Annu. Rev. Biochem.* **2005**, *74*, 681.

(22) Branzei, D.; Foiani, M. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (4), 297.

(23) Brown, J.; Brown, T.; Fox, K. R. *Biochemical Journal* **2001**, *354* (3), 627.

(24) Neidle, S. In *Principles of Nucleic Acid Structure*; Elsevier, 2008; pp 132–203.

(25) Jackson, B. A.; Barton, J. K. *J. Am. Chem. Soc.* **1997**, *119* (52), 12986.

(26) Jackson, B. A.; Alekseyev, V. Y.; Barton, J. K. *Biochemistry* **1999**, *38* (15), 4655.

(27) Erkkila, K. E.; Odom, D. T.; Barton, J. K. *Chem. Rev.* **1999**, *99* (9), 2777.

(28) Nakatani, K. *Bull. Chem. Soc. Jpn.* **2009**, *82* (9), 1055.

(29) Teulade-Fichou, M.-P.; Vigneron, J.-P.; Lehn, J.-M. *Supramolecular Chemistry* **1995**, *5* (2), 139.

(30) David, A.; Bleimling, N.; Beuck, C.; Lehn, J.-M.; Weinhold, E.; Teulade-Fichou, M.-P. *ChemBioChem* **2003**, *4* (12), 1326.

(31) Granzhan, A.; Largy, E.; Saettel, N.; Teulade-Fichou, M.-P. *Chem. Eur. J.* **2009**, *16* (3), 878.

(32) Mirkin, S. M. *Curr. Opin. Struct. Biol.* **2006**, *16* (3), 351.

(33) Kim, J. C.; Mirkin, S. M. *Current Opinion in Genetics & Development* **2013**, *23* (3), 280.

(34) McMurray, C. T. *Nat. Rev. Genet.* **2010**, *11* (11), 786.

(35) Pearson, C. E.; Edamura, K. N.; Cleary, J. D. *Nat. Rev. Genet.* **2005**, *6* (10), 729.

(36) Mirkin, S. M. *Nature* **2007**, *447* (7147), 932.

(37) Schoser, B.; Timchenko, L. T. *Curr. Genomics* **2010**, *11* (2), 77.

(38) Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J.-P.; Hudson, T.; Sohn, R.; Zemelman, B.; Snell, R. G.; Rundle, S. A.; Crow, S.; Davies, J.; Shelbourne, P.; Buxton, J.; Jones, C.; Juvonen, V.; Johnson, K.; Harper, P. S.; Shaw, D. J.; Housman, D. E. *Cell* **1992**, *68* (4), 799.

(39) Liquori, C. L.; Ricker, K.; Moseley, M. L.; Jacobsen, J. F.; Kress, W.; Naylor, S. L.; Day, J. W.; Ranum, L. P. W. *Science* **2001**, *293* (5531), 864.

(40) Richards, R. I. *Hum. Mol. Genet.* **2001**, *10* (20), 2187.

(41) La Spada, A. R.; Taylor, J. P. *Nat. Rev. Genet.* **2010**, *11* (4), 247.

(42) López Castel, A.; Cleary, J. D.; Pearson, C. E. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (3), 165.

(43) Liu, G.; Leffak, M. *Cell & bioscience* **2012**.

(44) Pearson, C. E.; Sinden, R. R. *Biochemistry* **1996**, *35* (15), 5041.

(45) Ohshima, K.; Wells, R. D. *J. Biol. Chem.* **1997**, *272* (27), 16798.

(46) Osborne, R. J.; Thornton, C. A. *Hum. Mol. Genet.* **2006**, *15 Spec No 2* (Review Issue 2), R162.

(47) Kuyumcu-Martinez, N. M.; Wang, G.-S.; Cooper, T. A. *Molecular Cell* **2007**, *28* (1), 68.

(48) Childs-Disney, J. L.; Stepniak-Konieczna, E.; Tran, T.; Yildirim, I.; Park, H.; Chen, C. Z.; Hoskins, J. W.; Southall, N.; Marugan, J. J.; Patnaik, S.; Zheng, W.; Austin, C. P.; Schatz, G. C.; Sobczak, K.; Thornton, C. A.; Disney, M. D. *Nat Commun* **2013**, *4*, 2044.

(49) Kalsotra, A.; Singh, R. K.; Gurha, P.; Ward, A. J.; Creighton, C. J.; Cooper, T. A. *Cell Reports* **2014**, *6* (2), 336.

(50) Moseley, M. L.; Zu, T.; Ikeda, Y.; Gao, W.; Mosemiller, A. K.; Daughters, R. S.; Chen, G.; Weatherspoon, M. R.; Clark, H. B.; Ebner, T. J.; Day, J. W.; Ranum, L. P. W. *Nature Genetics* **2006**, *38* (7), 758.

(51) Zu, T.; Gibbens, B.; Doty, N. S.; Gomes-Pereira, M.; Huguet, A.; Stone, M. D.; Margolis, J.; Peterson, M.; Markowski, T. W.; Ingram, M. A. C.; Nan, Z.; Forster, C.; Low, W. C.; Schoser, B.; Somia, N. V.; Clark, H. B.; Schmechel, S.; Bitterman, P. B.; Gourdon, G.; Swanson, M. S.; Moseley, M.; Ranum, L. P. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (1), 260.

(52) Kanadia, R. N.; Shin, J.; Yuan, Y.; Beattie, S. G.; Wheeler, T. M.; Thornton, C. A.; Swanson, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (31), 11748.

(53) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. *J. Am. Chem. Soc.* **2008**, *130* (48), 16254.

(54) Warf, M. B.; Nakamori, M.; Matthys, C. M.; Thornton, C. A.; Berglund, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (44), 18551.

(55) Coonrod, L. A.; Nakamori, M.; Wang, W.; Carrell, S.; Hilton, C. L.; Bodner, M. J.; Siboni, R. B.; Docter, A. G.; Haley, M. M.; Thornton, C. A.; Berglund, J. A. *ACS Chem. Biol.* **2013**, *8* (11), 2528.

(56) Hagihara, M.; Nakatani, K. *Nucleic Acids Symposium Series* **2006**, *50* (1), 147.

(57) Cinesi, C.; Aeschbach, L.; Yang, B.; Dion, V. *Nat Commun* **2016**, *7*, 13272.

(58) Gomes-Pereira, M.; Monckton, D. G. *Nucleic Acids Research* **2004**, *32* (9), 2865.

(59) Hashem, V. I.; Pytlos, M. J.; Klysik, E. A.; Tsuji, K.; Khajavi, M.; Khajav, M.; Ashizawa, T.; Sinden, R. R. *Nucleic Acids Research* **2004**, *32* (21), 6334.

(60) Yang, Z.; Lau, R.; Marcadier, J. L.; Chitayat, D.; Pearson, C. E. *Am. J. Hum. Genet.* **2003**, *73* (5), 1092.

(61) Scrudato, Lo, M.; Poulard, K.; Klein, A.; Tomé, S.; Martin, S.; Gourdon, G.; Furling, D.; Buj-Bello, A. *Neuromuscular Disorders* **2017**, *27*, S181.

(62) Arambula, J. F.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (38), 16068.

(63) Wong, C.-H.; Nguyen, L.; Peh, J.; Luu, L. M.; Sanchez, J. S.; Richardson, S. L.; Tuccinardi, T.; Tsoi, H.; Chan, W. Y.; Chan, H. Y. E.; Baranger, A. M.; Hergenrother, P. J.; Zimmerman, S. C. *J. Am. Chem. Soc.* **2014**, *136* (17), 6355.

(64) Jahromi, A. H.; Nguyen, L.; Fu, Y.; Miller, K. A.; Baranger, A. M.; Zimmerman, S. C. *ACS Chem. Biol.* **2013**, *8* (5), 1037.

(65) Jahromi, A. H.; Fu, Y.; Miller, K. A.; Nguyen, L. *J. Med. Chem.* **2013**, *56* (23), 9471.

(66) Wheeler, T. M.; Sobczak, K.; Lueck, J. D.; Osborne, R. J.; Lin, X.; Dirksen, R. T.; Thornton, C. A. *Science* **2009**, *325* (5938), 336.

(67) Parkesh, R.; Childs-Disney, J. L.; Nakamori, M.; Kumar, A.; Wang, E.; Wang, T.; Hoskins, J. W.; Tran, T.; Housman, D.; Thornton, C. A.; Disney, M. D. *J. Am. Chem. Soc.* **2012**, *134* (10), 4731.

(68) Garcia-Lopez, A.; Llamusí, B.; Orzáez, M.; Pérez-Payá, E.; Artero, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (29), 11866.

(69) Lian, C.; Robinson, H.; Wang, A. H.-J. *J. Am. Chem. Soc.* **1996**, *118* (37), 8791.

(70) Childs-Disney, J. L.; Hoskins, J. W.; Rzuczek, S. G.; Thornton, C. A.; Disney, M. D. *ACS Chem. Biol.* **2012**, *7* (5), 856.

(71) Disney, M. D.; Lee, M. M.; Pushechnikov, A.; Childs-Disney, J. L. *ChemBioChem* **2010**, *11* (3), 375.

(72) Lee, M. M.; Childs-Disney, J. L.; Pushechnikov, A.; French, J. M.; Sobczak, K.; Thornton, C. A.; Disney, M. D. *J. Am. Chem. Soc.* **2009**, *131* (47), 17464.

(73) Wong, C.-H.; Richardson, S. L.; Ho, Y.-J.; Lucas, A. M. H.; Tuccinardi, T.; Baranger, A. M.; Zimmerman, S. C. *ChemBioChem* **2012**, *13* (17), 2505.

(74) Marcheschi, R. J.; Tonelli, M.; Kumar, A.; Butcher, S. E. *ACS Chem. Biol.* **2011**, *6* (8), 857.

(75) Nguyen, L.; Luu, L. M.; Peng, S.; Serrano, J. F.; Chan, H. Y. E.; Zimmerman, S. C. *J. Am. Chem. Soc.* **2015**, *137* (44), 14180.

(76) Luu, L. M.; Nguyen, L.; Peng, S.; Lee, J.; Lee, H. Y.; Wong, C.-H.; Hergenrother, P. J.; Chan, H. Y. E.; Zimmerman, S. C. *ChemMedChem* **2016**, *11* (13), 1428.

(77) Mooers, B. H. M.; Logue, J. S.; Berglund, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (46), 16626.

(78) Chen, C. Z.; Sobczak, K.; Hoskins, J. W.; Southall, N.; Marugan, J. J.; Zheng, W.; Thornton, C. A.; Austin, C. P. *Anal Bioanal Chem* **2012**, *402* (5), 1889.

(79) Ho, Y.-J. LEAD COMPOUND DISCOVERY FOR MYOTONIC DYSTROPHY, 2013, pp 1–66.

(80) Mao, C.; Flavin, K. G.; Wang, S.; Dodson, R.; Ross, J.; Shapiro, D. J. *Analytical Biochemistry* **2006**, *350* (2), 222.

(81) Kuyper, L. F.; Baccanari, D. P.; Jones, M. L.; Hunter, R. N.; Tansik, R. L.; Joyner, S. S.; Boytos, C. M.; Rudolph, S. K.; Knick, V.; Wilson, H. R.; Caddell, J. M.; Friedman, H. S.; Comley, J.; Stables, J. N. *J. Med. Chem.* **1996**, *39* (4), 892.

(82) Guan, J.; Zhang, Q.; O'Neil, M.; Obaldia, N.; Ager, A.; Gerena, L.; Lin, A. J. *Antimicrob. Agents Chemother.* **2005**, *49* (12), 4928.

(83) Blandamer, M. J. *Biocalorimetry: Applications of Calorimetry in the Biological Sciences*; Ladbury, J. E., Chowdhry, B. Z., Eds.; Wiley: Chichester, 1998.

(84) Brown, A. *IJMS* **2009**, *10* (8), 3457.

(85) Freiburger, L.; Auclair, K.; Mittermaier, A. *Methods* **2015**, 1.

(86) Hunter, C. A.; Anderson, H. L. *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7488.

(87) Turnbull, W. B.; Daranas, A. H. *J. Am. Chem. Soc.* **2003**, *125* (48), 14859.

(88) Chen, J.; Kassenbrock, A.; Li, B. X.; Xiao, X. *Med. Chem. Commun.* **2013**, *4* (9), 1275.

(89) Roberts, R. J.; Cheng, X. *Annu. Rev. Biochem.* **1998**, *67* (1), 181.

(90) Roberts, R. J. *Cell* **1995**, *82* (1), 9.

(91) Priyakumar, U. D.; MacKerell, A. D. *Chem. Rev.* **2006**, *106* (2), 489.

(92) Pan, B.; Sundaralingam, M. *International Journal of Quantum Chemistry* **1999**, *75* (3), 275.

(93) Tubbs, J. L.; Latypov, V.; Kanugula, S.; Butt, A.; Melikishvili, M.; Kraehenbuehl, R.; Fleck, O.; Marriott, A.; Watson, A. J.; Verbeek, B.; McGown, G.; Thorncroft, M.; Santibanez-Koref, M. F.; Millington, C.; Arvai, A. S.; Kroeger, M. D.; Peterson, L. A.; Williams, D. M.; Fried, M. G.; Margison, G. P.; Pegg, A. E.; Tainer, J. A. *Nature* **2009**, *459* (7248), 808.

(94) Weidmann, A. G.; Komor, A. C.; Barton, J. K. *Comments on Inorganic Chemistry* **2014**, *34* (3-4), 114.

(95) Spies, M. A.; Schowen, R. L. *J. Am. Chem. Soc.* **2002**, *124* (47), 14049.

(96) Bando, T.; Fujimoto, J.; Minoshima, M.; Shinohara, K.-I.; Sasaki, S.; Kashiwazaki, G.; Mizumura, M.; Sugiyama, H. *Bioorganic & Medicinal Chemistry* **2007**, *15* (22), 6937.

(97) Shibata, T.; Nakatani, K. *ChemBioChem* **2016**, *17* (18), 1685.

(98) Wong, C.-H. DEVELOPING POTENTIAL DRUGS FOR THE TREATMENT OF MYOTONIC DYSTROPHY: FROM RATIONAL DESIGN TO LEAD DEVELOPMENT, 2013, pp 1–123.

(99) Arambula, J. F. PART I. A STACKED INTERCALATOR APPROACH TO SELECTIVELY TARGET MISMATCHES WITHIN CTG AND CUG TRINUCLEOTIDE REPEATS. PART II. AROMATIC STACKING OF NAPHTHOYL MODIFIED CYTOSINES WITHIN PNA•DNA DUPLEXES, 2010, pp 1–168.

(100) Constant, J. F.; Laûgaa, P.; Roques, B. P.; Lhomme, J. *Biochemistry* **1988**, *27* (11), 3997.

(101) Fkyerat, A.; Demeunynck, M.; Constant, J. F.; Michon, P.; Lhomme, J. *J. Am. Chem. Soc.* **1993**, *115* (22), 9952.

(102) Alarcon, K.; Demeunynck, M.; Lhomme, J.; Carrez, D.; Croisy, A. *Bioorganic & Medicinal Chemistry* **2001**, *9* (7), 1901.

(103) Wong, C.-H.; Fu, Y.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. *Nucleic Acids Research* **2011**, *39* (20), 8881.

(104) Leonard, N. J.; Lambert, R. F. *J. Org. Chem.* **1969**, *34* (11), 3240.

(105) Rewcastle, G. W.; Denny, W. A. *Synthetic Communications* **1987**, *17* (3), 309.

(106) Leonard, N. J. *Acc. Chem. Res.* **1979**, *12* (12), 423.

(107) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1999**, *38* (12), 3468.

(108) Rojsitthisak, P.; Romero, R. M.; Haworth, I. S. *Nucleic Acids Research* **2001**, *29* (22), 4716.

(109) Kaul, M.; Pilch, D. S. *Biochemistry* **2002**, *41* (24), 7695.

(110) Hayatsu, H.; Ukita, T. *Biochem. Biophys. Res. Commun.* **1967**, *29* (4), 556.

(111) Nakatani, K.; Sando, S.; Kumasawa, H.; Kikuchi, J.; Saito, I. *J. Am. Chem. Soc.* **2001**, *123* (50), 12650.

(112) Vögtle, F.; Pawlitzki, G.; Hahn, U. In *Modern Cyclophane Chemistry*; 2005; pp 41–80.

(113) Cram, D. J.; Cram, J. M. *Acc. Chem. Res.* **1971**, *4* (6), 204.

(114) Lehn, J.-M.; Schmidt, F.; Vigneron, J.-P. *Tetrahedron Letters* **1988**, *29* (41), 5255.

(115) Jazwinski, J.; Blacker, A. J.; Lehn, J.-M.; Cesario, M.; Guilhem, J.; Pascard, C. *Tetrahedron Letters* **1987**, *28* (48), 6060.

(116) Ramaiah, D.; Neelakandan, P. P.; Nair, A. K.; Avirah, R. R. *Chem. Soc. Rev.* **2010**, *39* (11), 4158.

(117) Zimmerman, S. C.; Lamberson, C. R.; Cory, M. *J. Am. Chem. Soc.* **1989**, *111* (17), 6805.

(118) Yang, X.-L.; Robinson, H.; Gao, Y.-G.; Wang, A. H.-J. *Biochemistry* **2000**, *39* (36), 10950.

(119) Jourdan, M.; Garcia, J.; Lhomme, J.; Teulade-Fichou, M.-P.; Vigneron, J.-P.; Lehn, J.-M. *Biochemistry* **1999**, *38* (43), 14205.

(120) Mukherjee, S.; Dohno, C.; Asano, K.; Nakatani, K. *Nucleic Acids Research* **2016**, gkw672.

(121) Mukherjee, S.; Dohno, C.; Nakatani, K. *Chem. Eur. J.* **2017**, *23* (47), 11385.

(122) Goodell, J. R.; Madhok, A. A.; Hiasa, H.; Ferguson, D. M. *Bioorganic & Medicinal Chemistry* **2006**, *14* (16), 5467.

(123) de Hoog, P.; Gamez, P.; Driessen, W. L.; Reedijk, J. *Tetrahedron Letters* **2002**, *43* (38), 6783.

(124) Baliani, A.; Bueno, G. J.; Stewart, M. L.; Yardley, V.; Brun, R.; Barrett, M. P.; Gilbert, I. H. *J. Med. Chem.* **2005**, *48* (17), 5570.

(125) Skehan, P.; Storeng, R.; Scudiero, D.; Monks, A.; McMahon, J.; Vistica, D.; Warren, J. T.; Bokesch, H.; Kenney, S.; Boyd, M. R. *J. Natl. Cancer Inst.* **1990**, *82* (13), 1107.

(126) Vichai, V.; Kirtikara, K. *Nat Protoc* **2006**, *1* (3), 1112.

(127) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. *Annu. Rev. Biochem.* **2008**, *77* (1), 383.

(128) Mizoue, L. S.; Tellinghuisen, J. *Analytical Biochemistry* **2004**, *326* (1), 125.